THE APPLICATION OF EXPONENTIAL SMOOTHING TO FORECASTING A TIME SERIES

by 884

RAYMOND CHARLES MILLER

B. S., University of Illinois, 1966

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1968

Approved by:

L. E. Grosh

Major Professor

TABLE OF CONTENTS

INTRODUCTION

A time series can be defined as a sequence of discrete observations taken at uniform intervals on a time scale. Since time series occur frequently in industrial and economic situations and since many decisions are based on the predictions of the future values of these time series it is an important decision aid to develop a method for forecasting these time series.

In order to treat the time series analytically it is described as the sum of two components. The first is the process which generates the series and the second is some superimposed random noise, or variation in the observations. Hence, the time series can be represented as

$$y(t) = \xi(t) + \varepsilon(t)$$

where $\xi(t)$ is the value of the process at time t and $\varepsilon(t)$ is the random noise associated with the observation at time t. The distribution of the noise samples has the following properties:

(1) The expected value is zero, i.e.

$$E\big(\varepsilon(t)\big) = 0$$

(2) The noise samples have no series correlation, i.e.

$$E\big(\varepsilon(i)\varepsilon(j)\big) = 0 \quad \text{for } i \neq j$$

Considering the nature of the time series the object of the forecasting technique is to provide a forecast that will

(1) Dampen, or smooth, the effect of the random noise,

(2) Reflect any trends which the time series might be undergoing, i.e. the time series need not be stationary,

and

(3) Provide an unbiased estimate of the time series.

Hence, a forecasting technique is sought which will seek a balance between the response to secular trends in the process and the errors caused by the random noise.

The process which generates the time series can be described by two components. These are the secular trend of the mean and the periodic component of the series. Along with the trend which the mean of the time series might be undergoing the periodic component might also contain trends. These latter trends are of two types

(1) Changing amplitudes

(2) Shifting phase angles

Therefore, the forecasting model must adequately represent the time series and be able to adjust to these trends. It is found that forecasting models composed of polynomials, trigonometric functions and multiples of these functions fulfill the requirements of the forecasting model.

These forecasting models can be represented in vector form as

$$y(t) = a'f(t)$$

where a is a vector of coefficients and f(t) the vector of fitting functions evaluated at time t. The requirements of the forecast are satisfied in the method by which the vector of coefficients is estimated. The estimation of this vector is performed by the use of general exponential smoothing. In this process the estimate of the coefficients is based on the discounted least squares criterion. That is, the sum

$$\sum_{j=0}^{\infty} \beta^j \left(y(-t) - \hat{y}(-t-1)\right)$$

is minimized where y(t) is the value of the time series at time t,

$\hat{y}(t-1)$ is the forecast of this value at the previous period,

β is a positive constant less than one.

This forecasting technique has the following properties:

(1) The weight given the observations in the forecast is discounted on a time scale at a rate which can be controlled by the value of the discount factor β, thereby yielding control of the responsiveness of the forecast.

(2) All past data is contained in a single word of information.

(3) A simple recursive relationship can be developed for re-evaluating the estimates of the vector of coefficients with each observation.

The justification for the use of general exponential smoothing lies in the proof that if the forecasting model is a true representation of the process then the expected value of the vector of coefficients obtained by exponential smoothing will be the true value of this vector. The problem, then, becomes that of determining the forecasting model. It is found that the trend of the mean can be represented by a polynomial. To this end polynomial regression is used to fit the curve representing the trend of the mean to the data. Having performed this regression both qualitative and quantitative estimates for the terms in the forecasting model which represent the trend of the mean are available. The regression curve is then subtracted from the data, an operation referred to as "detrending".

Making the data available in the "detrended" form is the first phase in the analysis of the periodic component. Essentially "detrending" allows the periodic component to be observed seperately. The second analysis performed on the periodic component is autocorrelation analysis. The autocorrelation function is defined and shown to have a maximum at the basic

period of the data. Moreover, autocorrelation analysis can be used as a test of significance that a process exists.

The basic theorem of Fourier series states that if a discrete series is basically periodic then the coefficients of a series of sines and cosines of the fundamental harmonics of that period can be determined so that the Fourier series yields the data points. This analysis is extended to a trigonometric series which includes all the frequencies in the data and then to a least squares evaluation of these coefficients. It is shown that a measure of the contribution of each frequency in describing the time series can be expressed as a function of these coefficients and that using this measure the periodic component can be adequately expressed as a limited number of these frequencies.

A study is made of the sensitivity of the forecasts obtained through the use of exponential smoothing. The sensitivity analysis is carried out on the following parameters

(1) The fitting functions used to describe the forecasting model,

(2) The basic period of the forecasting model, and

(3) The value of the discount factor.

It is found that the choice of each of these parameters significantly affects the accuracy of the forecasts. In the case of the smoothing constant a literature search reveals that no method of determining the optimal value of this constant has been presented. This research does not attempt to develop such a method. It is felt, however, that a local optimal value for the discount factor might exist for a particular time series and set of fitting functions. Hence, a parametric investigation of the smoothing constant is carried out.

A general program to perform exponential smoothing was written. This program has the ability to

(1) Change the forecasting modcl,

(2) Change the basic period of the forecasting model,

(3) Change the value of the smoothing constant and

(4) Change the time series.

This program has the dual purpose of serving as a medium for carrying out and evaluating the effectiveness of general exponential smoothing for forecasting a time series and determining the feasibility and economy of performing this forecasting technique using a Fortran type of processor. Moreover, the results can be compared with those obtained by Brown (1) to verify the consistancy and accuracy of the program. The general nature of the program is necessary in order to carry out the sensitivity analysis on the forecast parameters and the parametric investigation of the smoothing constant. Furthermore this type of general program can serve to illustrate that the requirements of an industrial situation wherein many and varied time series are encountered can be satisfied.

Using the I.B.M. 1410 system at Kansas State University the program for general exponential smoothing can not be accomodated in the available core storage capacity. Hence, phasing of the program is necessary. The system at Kansas State University is programmed internally with PR-155 and has seven magnetic tape drive units which makes phasing possible. Each phase is run independently and upon its completion the processor automatically clears core and loads in the next phase. Any information required for following phases is written onto a work tape. In this system one work tape is availablc, however, after the program is compiled two more of the

tapes can be used as scratch files. These three tape units are the minimum required to perform the internal data transmission between phases.

Two time series are used in the application of the forecasting techniques. The first is the number of miles traveled on international airline routes measured at monthly intervals from January 1949 to December 1960. This data was obtained from Smoothing Forecasting and Prediction of Discrete Time Series by Robert Goodell Brown. The second time series used was the sheep population in England and Wales measured in yearly intervals from 1867 to 1939. This data was obtained from The Advanced Theory of Statistics Volume II by M. G. Kendall.

1. REPRESENTATION OF THE TIME SERIES

1.1 The Time Series Model

A time series is a sequence of observations taken at equal time intervals. For the purpose of analysis the time series can be considered to be made up of two elements

(1) The process which generates the time series

(2) Some superimposed random noise

Thus the time series may be represented on the following manner

$$y(t) = \xi(t) + \varepsilon(t)$$

where $\xi(t)$ is the process

$\varepsilon(t)$ is the noise in the t^{th} observation.

The distribution of $\varepsilon(t)$ has the properties

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t \varepsilon_j) = 0 \quad \text{for } i \neq j$$

$$= \sigma^2 \text{ for } i = j$$

where σ^2 is the variance of the noise distribution.

Recognizing that the noise in the time series is random, no attempt is made to represent it. The techniques of representing the time series are concerned with the process. Due to the conditions which generate time series it is generally adequate to describe the process in terms of two components

(1) The trend which the mean of the series is following,

(2) A cyclical component which is superimposed upon this trend.

The trend component may be represented by the following functions

(1) Polynomials,

(2) Exponentials,

whereas the periodic component must necessarily be represented by trigonometric functions. This representation is based on the Fourier analysis which deals with functions, either continuous or discrete, by means of a series of fundamental harmonics. The principal theorem of Fourier series may be stated as follows:

> If $f(t)$ is a single-valued function which has a derivative throughout the interval $-a \leq t \leq a$ except for a finite number of points at which it has finite discontinuities, and for other values of t is defined by the equation
>
> $$f(t + 2a) = f(t)$$
>
> then $f(t)$ can be represented by means of the Fourier series

$$\begin{aligned} y = \tfrac{1}{2}A_0 + A_1 \cos(\Pi t/a) + A_2 \cos(2\Pi t/a) &+ \\ + A_3 \cos(3\Pi t/a) &+ \dots \\ + B_1 \sin(\Pi t/a) + B_2 \sin(2\Pi t/a)& \\ + B_3 \sin(3\Pi t/a) &+ \dots \end{aligned} \tag{1:1:1}$$

See appendix A for the development of the Fourier analysis.

In the application of the Fourier series to representing the periodic part of the time series only the terms in (1:1:1) which are shown to be significant will be used. A limited number of terms are necessary to strike a balance between the accuracy of the model and the length of the computations. Hence a general representation of the time series is:

$$\begin{aligned} y = \text{trend} &+ A(T) \cos(2\Pi t/T) + B(T) \sin(2\Pi t/T) \\ &+ A(T') \cos(2\Pi t/T') + B(T') \sin(2\Pi t/T') \end{aligned} \tag{1:1:2}$$

where T and T' are the periods of the harmonics which show significant contribution to the representation of the time series.

The general model, (1:1:2), is adequate for representing a time series which displays a secular trend in the mean and a periodic component superimposed on that trend. In some time series, however, two other types of trends may appear in the periodic component

(1) Shifting phase angles,

(2) Growing amplitudes.

The technique for representing these trends is to include in the model a set of terms of the form

$$(a_1 + a_2 t)\cos(2\Pi t/T) + (a_3 + a_4 t)\sin(2\Pi t/T) \tag{1:1:3}$$

where T is the harmonic in which either or both of the above trends occur.

An example of the representation of a time series might be sighted as follows. Consider the time series in figure (1:1:1)

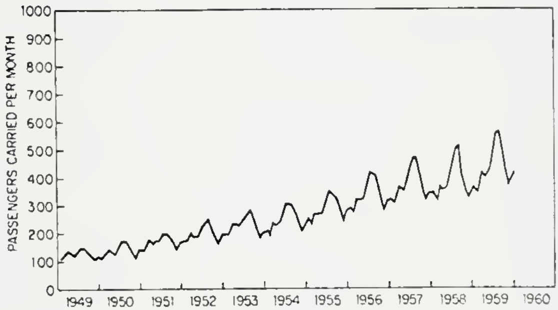

Fig. 1:1:1 International Airline Passenger Data

Looking ahead in the analysis momentarily this time series might be adequately represented by a linear trend with a periodic component of 12 months. Moreover, and without any justification at this point, a harmonic at 6 months

might be significant. Furthermore, the amplitude of the periodic component seems to be increasing. With these ideas the terms included in the model would be:

(1) $a_0 + a_1 t$ to represent the linear trend,

(2) $(a_2 + a_4 t) \sin(2\Pi t/12) + (a_3 + a_5 t) \cos(2\Pi t/12)$

to represent the 12 month periodic component with increasing amplitude,

(3) $a_6 \sin(4\Pi t/12) + a_7 \cos(4\Pi t/12)$

to represent the harmonic.

1:2 Trend Analysis of a Time Series

The previous section showed that an adequate representation of the time series depended on proper evaluation of the trends that the series was following. These trends are of two types:

(1) Trend of the mean

(2) Trend of the periodic component.

Moreover, the analysis of the periodic component was treated independently of the time series. In effect this independent treatment is equivalent to evaluating the series with the trend of the mean removed. The purpose of trend analysis of the time series, then, is twofold.

(1) Trend analysis provides a quantitative estimate of the trends in the time series.

(2) Trend analysis removes the trend of the mean from the time series. This "detrended" form of the data is used in further analysis.

The mechanism used for trend analysis is polynomial regression. (see appendix B). In the expression for time time series

$$y = \xi(t) + \varepsilon(t)$$

the process $\xi(t)$ can be written as

$$\xi(t) = \text{trend} + \text{periodic component}$$

If the trend component is expressed as y'' then the expression for the time series with the trend removed is

$$\begin{aligned} y' &= \xi(t) + \varepsilon(t) - y'' \\ &= y - y'' \end{aligned} \tag{1.2.1}$$

where y'' is the value of the regression curve. Hence y' is the periodic component of the time series.

In selecting the regression curve to represent the trend of the mean, care must be taken not to include any of the periodic component in the trend to be removed. The trigonometric functions can be expressed as a series

$$\begin{aligned} \sin t &= t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} + \ldots \\ \cos t &= 1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \frac{t^6}{6!} + \ldots \end{aligned} \tag{1.2.2}$$

Now suppose the regression curve is taken to be a polynomial of degree k.

$$y = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k \tag{1.2.3}$$

If, in fact, the time series consists of a trend component which can be expressed as a polynomial of degree k', where $k' < k$, and a periodic component which can be expressed as

$$y' = A(T)\cos(2\pi t/T) + B(T)\sin(2\pi t/T)$$

then considering (1:2:2) the series y can be expressed as a polynomial of degree k", where $k'' > k$.

$$y = a_0 + a_1 t + a_2 t^2 + \ldots + a_{k''} t^{k''}$$

Hence if the order of the regression takes on the value k, where $k > k'$ then the regression curve obtained can be expressed as the sum of two series

$$y''_\alpha = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_k t^k$$

$$y''_\beta = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_k t^k$$

where y_α is the series contributed by the secular trend of the mean and y_β is the series contributed by the periodic component. Hence, the value obtained for the regression curve will be

$$y'' = y''_\alpha + y''_\beta$$

where in fact it should be

$$y'' = y''_\alpha$$

Therefore if $k > k'$ then part of the periodic component will be removed from the data in "detrending" and the trend analysis will give a distorted picture of the data. That is the "detrended" data will be

$$y' = y - (y_\alpha + y_\beta)$$

instead of

$$y' = y - y_\alpha \ .$$

An example of trend analysis can be found in its application to the time series in figure (1:1:1). In this case the secular trend of the mean is taken to be a linear relationship. That is the expression for the trend of the mean is taken to be

$$y'' = a_0 + a_1 t$$

where a_1 and a_0 are the coefficients to be determined by regression analysis. Appendix B gives the expressions for the least squares estimate of these coefficients as

$$a_0 = \bar{y}$$

$$a_1 = \frac{\sum_{i=1}^{n} ty - \bar{t} \sum_{i=1}^{n} y}{\sum_{i=1}^{n} t^2 - 2\bar{t} \sum_{i=1}^{n} t + \sum_{i=1}^{n} (\bar{t})^2}$$

Hence the expression for the detrended time series is

$$y'(t) = y(t) - t\, a_1$$

where t denotes the t^{th} value in the time series

Thus far the trend analysis has fulfilled the purposes of providing the time series in detrended form and providing quantative estimates for the secular trend in the mean. A third purpose of trend analysis is that of providing quantitative estimates for the trends of the periodic component of the time series. With the time series expressed in the form

$$y' = y - y''$$

the periodic component can be looked at directly without any effect of the secular trend of the mean. Thus if the amplitude of the periodic component is undergoing a secular trend the series would appear in the detrended form as

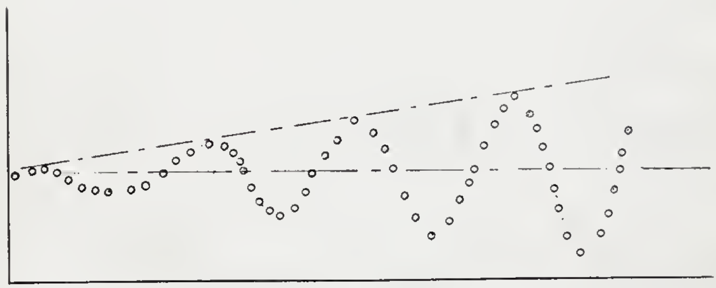

Fig. 1:2:1 Example of "detrended" time series

in which case the slope of the line A'A' can be easily determined. If however, the mean is following a secular trend which could be expressed as

$$y'' = a_0 + a_1 t + a_2 t^2$$

the series before trend analysis is applied might look like figure (1:2:2). Hence, the preceeding analysis would be much more difficult to perform.

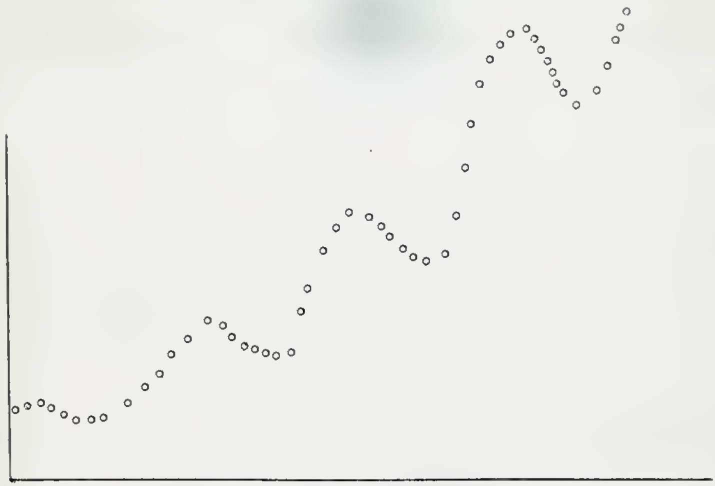

Fig. 1:2:2 Series Before "Detrending"

1.3 Autocorrelation Analysis of the Time Series

In the previous section the time series was expressed as

$$y(t) = \xi(t) + \varepsilon(t)$$

where $\xi(t)$ is the process

$\varepsilon(t)$ noise samples, where $\varepsilon(t)$ is the t^{th} sample from a probability distribution with zero mean

The application of the principal theorem of Fourier series is based on the assumption that a periodic component of the time series exists which can be represented as

$$f(t + 2a) = f(t)$$

The purpose of autocorrelation analysis is to answer three questions about the time series

(1) Does a process exist?

(2) Does a periodic component exist?

(3) What is the basic period of the time series?

The mechanism used in answering these questions, correlation analysis, is developed in Appendix C.

Consider the term

$$\sum_{j=k+1}^{T} y_j y_{j-k} \tag{1.3.1}$$

where y_j is the j^{th} observation in the time series

$j = 1,2,\ldots, T$

k is the lag between the terms in the summation

The expected value of this term is denoted as the average lagged product. In particular, if the sequence y has been adjusted so that the expected value, $E(y) = 0$, then the average lagged product is

$$R_{xx}(k) = \frac{\sum_{j=k+1}^{T} y_j y_{j-k}}{T-1-k} \tag{1.3.2}$$

the autocovariance.

The variance of a sequence of numbers that have been adjusted to have a mean value of zero is just the expected value of the squares of these numbers. This can be expressed as $E(y^2)$ which is effectively $R_{xx}(0)$. Moreover, the autocovariance expressed in normalized form

$$\rho(k) = R_{xx}(k)/R_{xx}(0) \tag{1.3.4}$$

is merely the autocorrelation coefficient. The set of values for the autocorrelation coefficient for all lags,

$k = 1, 2, \ldots\ldots$

is defined as the autocorrelation function. Hence, the autocorrelation function for a lag k=0 is one. Three other properties of the autocorrelation function are of significance in the analysis.

(1) The range of ρ is between $\pm$ 1.

(2) Pure random noise will have zero correlation between samples not identically equal to each other.

(3) If whenever y_t is positive so is y_{t+k}. And whenever y_t is negative so is y_{t+k}. Then the autocovariance will be large and positive. In this case, pairs of observations k units of time apart in the sequence are highly correlated and one can be used to forecast the other. In a similiar manner if y_t positive usually implies that y_{t+k} is negative (and vice versa), one can still be used to forecast the other. The autocovariance in this case will be large and negative (1,394).

In the application of autocorrelation analysis Brown (1,395) suggests the following proceedure. Plot the observations to see if you should expect a secular trend or a significant cyclical pattern. If there is a secular trend, fit a straight line to the data by least squares. Using this least squares fit adjust the data to zero expected value. Compute the autocovariances using (1:3:2).

$$R_{xx}(k) = (1/T-1-k) \sum_{j=k+1}^{T} y_j y_{j-k}$$

This suggested method of Brown's can be applied more discretly by taking advantage of the trend analysis in the previous section. In that section the data was obtained in the detrended form as

$y' = y - y''$

where y is the time series

y'' is the secular trend of the mean

Moreover, since the time series can be expressed as

y = trend + periodic component + random noise

it follows that

y' = periodic component + random noise

It was shown in appendix A that by the principal theorem of Fourier series the periodic component can be represented by the series

$$y' = \tfrac{1}{2}A_0 + A_1 \cos(\Pi t/a) + A_2 \cos(2\Pi t/a) + A_3 \cos(3\Pi t/a) + \dots$$
$$+ B_1 \sin(\Pi t/a) + B_2 \sin(2\Pi t/a) + B_3 \sin(3\Pi t/a) + \dots$$

Note that the expected value of this series is merely $\tfrac{1}{2}A_0$. If this expected value is subtracted from the "detrended" series the following is obtained.

$$\hat{y} = y' - \tfrac{1}{2}A_0 + e(t)$$

where $e(t)$ is the noise sample

The expected value of this series can be expressed as

$$E(\hat{y}) = E(y' - \tfrac{1}{2}A_0) + E(e(t))$$

Now since the expected value of the first term on the right was shown to be zero and since by definition the expected valuc of the noise is zero, then

$$E(\hat{y}(t)) = 0$$

and this series satisfies the conditions for the application of (1.3.2), the autocovariance.

To facilitate the use of the autocorrelation function in answering the question previously posed in this section, i.e.

(1) Does a process exist?

(2) Does a periodic component exist?

(3) What is the basic period of the series?

some of the properties of this function must be noted.

An example of autocorrelation analysis is given bt H. T. Davis

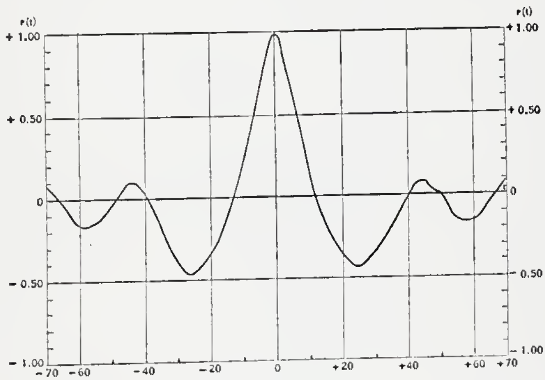

Figure 1:3:1 Autocorrelation function for industrial stock prices, t measured in months

Referring to this plot Davis comments:

> '"It will be observed from the graph that the function damps rapidly. It changes from positive to negative at approximately $t = \pm 10$, and again becomes positive at $t = \pm 40$. As we shall show later on this may be interpreted as indicating a cycle of 40 months." (3.356)

Hence the autocorrelation function will reach its maximum at the basic period of the series. The above characteristic may be proved as follows.

The autocovariance (1.3.2) may be expressed as

$$R_{xx}(k) = \lim_{t \to \infty} \frac{1}{2T + 1} \sum_{t=-T}^{T} y(t+k)\, y(t)$$

where the series y(t), t = 1,2,... T, has been adjusted so that the expected value is zero. But the series adjusted in this manner can be expressed as

$$y'(t) = y(t) - y''(t)$$

where y"(t) is the secular trend of the mean. Therefore, the adjusted series can be expressed as

$$y'(t) = \text{periodic component} + \text{noise}$$

In Appendix A it is shown that the periodic component can be expressed as a Fourier series

$$y'(t) = a_0 + \sum_{i=1}^{n} a_i \cos \omega_i + \sum_{i=1}^{n} b_i \cos \omega_i$$

Making the transformation from the general formulation of the Fourier series above to an infinite series of cosines (this change will only simplify the calculations), the autocovariance may be re-expressed as

$$R_{xx}(k) = \lim_{T} \frac{1}{2T+1} \sum_{t=-T}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \cos\left(\omega_i t\right) \cos\left(\omega_i (t+k)\right) + \varepsilon_t \varepsilon_{t+k}$$

since the cross product between the noise and the cosine signal have expected value zero. The expected value of all terms of the form

$$\cos(\omega_i t) \cos(\omega_j t)$$

is also zero for $i \neq j$. Therefore, the autocovariance reduces to

$$R_{xx}(k) = \frac{1}{2} \sum_{i=1}^{n} a_i^2 \cos \omega_i k + R_{\varepsilon\varepsilon}(k)$$

If the assumption is made that the noise has no serial correlation then

$$R_{\varepsilon\varepsilon}(k) = \sigma_\varepsilon^2 \delta(k)$$

and the autocorrelation function $\rho(k) = \frac{R_{xx}(k)}{R_{xx}(0)}$ will have a local maximum at $R_i = (2\Pi/\omega_i)$. This completes the proof. (1,396)

The autocorrelation function can be used to determine if a process exists. Davis (5,143) provides an example of the autocorrelation function for a completely random series. The random series was constructed in the following manner. The percentages of trend of the Dow-Jones industrial averages for the prewar period (1897-1913) were written on cards and these cards were drawn at random to form a series of 204 items, that is N = 204. The polt of the autocorrelation function was determined to be:

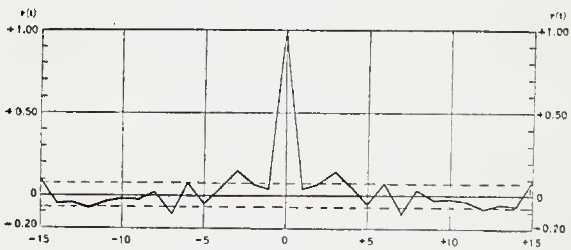

Fig. 1:3:2 Autocorrelation of a random series- the dotted lines define the standard crror band

The standard error band is computed by the relationship

$$S^2_{yx} = \frac{N-1}{N-2} (S^2_y - b^2 S^2_x)$$

where b is given by the regression formula (5)

$$b = \frac{\sum x_i y_i - \sum x_i \sum (y_i)/N}{x_i^2 - (\sum x_i)^2/N} .$$

The standard error band varies from ± 0.070 at the beginning where N = 204, to $\pm .0076$ at the end, where N = 204 - 30 = 174. Hence the distribution of the lagged values is seen to meet the test of randomness in a satisfactory manner.

1:4 Spectral Analysis of the Time Series

The original definition of the time series took the form

$$y = \xi(t) + \varepsilon(t)$$

where $\xi(t)$ is the process

$\varepsilon(t)$ is the random noise

Moreover, the process $\xi(t)$ can be expressed as

$$\xi(t) = \text{trend} + \text{periodic component}$$

The purpose of this section is to investigate the periodic component of the series.

The principal theorem of Fourier series (appendix A) shows that the periodic component of the time series can be represented to any desired accuracy by the series

$$y' = \tfrac{1}{2}A_0 + A_1 \cos(\Pi t/a) + A_2 \cos(2\Pi t/a) + A_3 \cos(3\Pi t/a) + \ldots$$
$$+ B_1 \sin(\Pi t/a) + B_2 \sin(2\Pi t/a) + B_3 \sin(3\Pi t/a) + \ldots \tag{1:4:1}$$

The terms of the above series represent the harmonics of the basic period of the time series. In the general case the process can be represented more accurately by increasing the number of terms in this series. However, these harmonic terms together with the terms which represent the secular trend of the mean form the forecasting model. Hence, as the number of terms in the Fourier series increases the time series is discribed more accurately but the calculations for the forecast are also increased.

The study of harmonic analysis shows that the frequencies represented in (1:4:1) differ in their contribution to representing the time series. The purpose of the spectral analysis is to obtain a measure of the contribution of each frequency. This measure is used as a basis for selecting the frequencies of the periodic terms to include in the forecasting model.

The analysis of the representation of the periodic component of the time series by the Fourier series (1:4:1) is based on the variance of the approximation. In this analysis the inequality of Bessel is used to express the variance of (1:4:1) in terms of the Fourier coefficients.

In order to derive this inequality assume that the process $y'(t)$ has been approximated by the first N harmonics of the Fourier series (1:4:1), that is

$$y'(t) \cong \tfrac{1}{2}A_0 + \sum_{n=1}^{N} A_n \cos(n\Pi t/a) + \sum_{n=1}^{N} B_n \sin(n\Pi t/a) \tag{1:4:2}$$

If the right-hand member of (1:4:2) is represented by $y_n(t)$ and the intergal of the square of the residual is considered, then

$$I = \frac{1}{a} \int_{-a}^{a} \left(y'(t) - y_n(t)\right)^2 dt$$

By expansion

$$= \frac{1}{a} \int_{-a}^{a} \left(y'^2(t) - 2y'(t)y_n(t) + y_n^2(t)\right) dt$$

Taking account of the integrals

$$\int_{-a}^{a} \sin(m\Pi t/a)\, \sin(n\Pi t/a)\, dt = \int_{-a}^{a} \cos(m\Pi t/a)\, \cos(n\Pi t/a) = 0$$

for $m \neq n$

$$\frac{1}{a} \int_{-a}^{a} \sin^2(n\Pi t/a)\, dt = \frac{1}{a} \int_{-a}^{a} \cos^2(n\Pi t/a)\, dt = 1$$

$$\int_{-a}^{a} \sin(m\Pi t/a)\, \cos(n\Pi t/a)\, dt = 0,$$

and observing the definitions of the Fourier coefficients given in Appendix A. The expression for the integral I can be obtained as

$$I = \frac{1}{a} \int_{-a}^{a} y'^2(t)\, dt - (\tfrac{1}{2}A_0^2 + R_1^2 + R_2^2 + \ . \ . \ . + R_N^2)$$

$$\text{where } R_n^2 = A_n^2 + B_n^2$$

Moreover, since the integrand of the integral is positive or zero the integral itself is positive or zero, and thus the Bessel inequality for Fourier coefficients is obtained. (3,65)

$$\tfrac{1}{2}A_0^2 + R_1^2 + R_2^2 + \ldots + R_N^2 \leq \frac{1}{a}\int_{-a}^{a}\left(f(t)\right)^2\,dt \qquad (1:4:3)$$

By noting that the arithmetic average of $y'(t)$ is equal to $\tfrac{1}{2}A_0$, the variance σ^2, of the residual function

$$\Delta t = y'(t) - y_n(t)$$

is given by

$$\sigma = \tfrac{1}{2}I = \tfrac{1}{2}\sum_{n=1}^{\infty}(R_n^2) = \tfrac{1}{2}\sum_{n=1}^{\infty}(A_n^2 + B_n^2) \qquad (1:4:4)$$

The term which is used as a measure of the contribution of each harmonic is

$$E(T) = \frac{R^2(T)}{2\sigma^2} \qquad (1:4:5)$$

where $R^2(T) = A^2(T) + B^2(T)$

T is the period of the harmonic

σ^2 is the variance of the data.

From Bessel's theorem, the variance σ_1^2, of the series after n terms have been removed; that is, equivalently, if the series were corrected for these harmonics, is (3,71)

$$\sigma_1^2 = (1 - \sum E_n)\sigma^2$$

Hence, the energies of the harmonics as expressed by (1:4:5) are strictly additive if the harmonics belong to the Fourier sequence. If the harmonics do not belong to the Fourier sequence then this expression is only approximately correct. Hence, $E(T)$ is an appropriate measure of the contribution of a harmonic.

The following expressions for the Fourier coefficients for the case of a discrete series are given in Appendix A.

$$A_n = \frac{2}{T} \sum_{t=1}^{T} f_t \cos(2n\Pi t/T)$$

$$B_n = \frac{2}{T} \sum_{t=1}^{T} f_t \sin(2n\Pi t/T) \qquad (1.4.6)$$

where T is the number of observations in the period of the harmonic

f_t is the time series evaluated at time t.

Upon consideration of these formulas a major objection is noted. The frequencies evaluated by equations (1.4.6) reflect only those frequencies which are present in the Fourier sequence. Since the purpose of harmonic analysis is to measure the contribution of all frequencies present, another approach must be considered.

To facilitate a more through approach to the harmonic analysis reference is made to an analysis by H.S. Carslaw under the topic of practical harmonic analysis and periodogram analysis. The objection sighted above was overcome by substituting for the Fourier series a trigonometric series with a limited number of terms. This is done in the following manner. Having the values of the time series for one period given at the points

$0, a, 2a, \ldots, (m-1)a$

where $ma = 2\Pi$

the equidistant points on the time axis at which the observations are taken are denoted by

$$x_0, x_1, \ldots, x_{m-1}$$

and the corresponding values of the observations by

$$y_0, y_1, y_2, \ldots, y_{m-1}$$

The time series is then represented by the sequence

$$f_n(x) = a_0 + a_1 \cos x_1 + a_2 \cos 2x_1 + \ldots + a_n \cos nx_1$$
$$+ b_1 \sin x_1 + b_2 \sin 2x_1 + \ldots + b_n \sin nx_1$$

If 2n+1=m the fourier coefficients can be determined so that

$$f(x_r) = y_r \quad \text{when } r = 0,1,2,\ldots,2n$$

The 2n+1 equations which yield this determination are

$$\begin{array}{ll}
a_0 + a_1 + \ldots + a_p + \ldots + a_n & = y_0 \\
a_0 + a_1 \cos x_1 + \ldots + a_p \cos px_1 + \ldots + a_n \cos nx_1 \\
\quad + b_1 \sin x_1 + \ldots + b_p \sin px_1 + \ldots + b_n \sin nx_1 & = y_1 \\
\ldots \\
a_0 + a_1 \cos x_{2n} + \ldots + a_p \cos px_{2n} + \ldots + a_n \cos nx_{2n} \\
\quad + b_1 \sin x_{2n} + \ldots + b_p \sin px_{2n} + \ldots + b_n \sin nx_{2n} & = y_{2n}
\end{array}$$

By adding the above equations

$$(2n+1)a_0 = \sum_{r=0}^{2n} y_r$$

since $1 + \cos pa + \cos 2pa + \ldots + \cos 2npa = 0$

and $\sin pa + \sin 2pa + \ldots + \sin 2npa = 0$

when $(2n + 1) = 2\Pi$.

Furthermore

$$1 + \cos(pa)\cos(ra) + \cos(2pa)\cos(2ra) + \ldots + \cos(2npa)\cos(2nra) = 0 \qquad p \neq r$$

$$\cos(pa)\sin(ra) + \cos(2pa)\sin(2ra) + \ldots + \cos(2npa)\sin(2nra) = 0 \qquad \begin{array}{l} p=1,2,\ldots n \\ r=1,2,\ldots n \end{array}$$

and

$$1 = \cos^2(pa) + \cos^2(2pa)+\ldots+ \cos^2(2npa) = \tfrac{1}{2}(2n+1)$$

Hence, if the second equation is multiplied by $\cos(px_1)$, the third by $\cos(px_2)$, etc. and added the result will be

$$\tfrac{1}{2}(2n+1)a_p = \sum_{r=0}^{2n} y_r \cos(pra)$$

In a similiar manner it can be determined that

$$\tfrac{1}{2}(2n+1)b_p = \sum_{r=1}^{2n} y_r \sin(pra)$$

Hence, a trigonometric series is formed whose sum takes the required values at the points

$$0,\ a,\ 2a,\ \ldots,\ 2na \quad \text{where } (2n+1)a = 2\Pi$$

If a period contains an even number of observations the relationships take the following forms. The interval of one period is denoted by

$$0,\ a,\ 2a,\ \ldots,\ (2n-1)a, \quad \text{where } na = 2\Pi$$

and the corresponding values of the observations are

$$y_0,\ y_1,\ y_2, \ldots\ y_{2n+1}$$

In this case the values of the 2n constants in the Fourier series

$$\begin{aligned} f_n(x) = a_0 + a_1\cos(x) + a_2\cos(2x) + \ldots + a_{n-1}\cos(n-1)x \\ + a_n\cos(nx) \\ + b_1\sin(x) + b_2\sin(2x) + \ldots + b_{n-1}\sin(n-1)x \end{aligned}$$

So that this series yields the points in the time series are (9)

$$a_0 = \frac{1}{2n} \sum_{r=0}^{2n+1} y_r$$

$$a_p = \frac{1}{n} \sum_{r=0}^{2n-1} y_r \cos pra, \text{ if } p \neq n$$

$$a_n = \frac{1}{2n} \sum_{r=0}^{2n-1} y_r \cos r\Pi$$

$$b_p = \frac{1}{n} \sum_{r=1}^{2n-1} y_r \sin pra$$

where $a = \Pi/n$

For the purpose of time series analysis the above equations while better than (1.4.6) are still inadequate. The expressions above depend on the criterion that the number of observations in the period is equal to the number of Fourier coefficients to be determined. In the application of Fourier analysis to the time series it is often desirable to determine the Fourier coefficients in a series in which $m > 2n+1$. That is, the number of coefficients in the series is less than the number of observations in the basic period. There are two justifications for the above condition

(1) The computation time for considering all m frequencies in the Fourier series may not be justifiable.

(2) The forecasting technique used involves matrix manipulation. Since the calculation time for matrix manipulation increases exponentially with the order of the matrix, and since the order of the matricies increases by 2 with each periodic term added to the forecasting model, these terms must necessarily be limited.

For the purpose of determining the Fourier coefficients for a series in which $m > 2n+1$ the previous analysis which depends on $2n+1 = m$ is, of course, no longer applicable. The result is a movement from the deterministic evaluation of the coefficients to an approximation of the coefficients based on the theory of least squares. That is, the coefficients

$$a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$$

are determined so that $f_n(x)$ approximates as closely as possible

$$y_0, y_1, \ldots, y_{m+1} \text{ at } x_0, x_1, \ldots, x_{m+1}$$

The theory of least squares shows that the closest approximation is obtained by making the function

$$\sum_{r=0}^{m-1}(y_r - f_n(x_r))^2$$

a minimum. Where the above sum is regarded as a function of

$$a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$$

The conditions which make this sum a minimum are given by Carslaw (2)

$$\sum_{r=0}^{m-1}(y_r - f_n(x_r)) = 0$$

$$\sum_{r=0}^{m-1}(y_r - f_n(x_r)) \cos px_r = 0$$

$$\sum_{r=0}^{m-1}(y_r - f_n(x_r)) \sin px_r = 0$$

where the above expressions are evaluated for $p = 1, 2, \ldots n$. The conditions above lead to the following values for the coefficients

$$ma_0 = \sum_{r=0}^{m-1} y_r$$

$$\tfrac{1}{2}ma_p = \sum_{r=0}^{m-1} y_r \cos prx_p$$

$$\tfrac{1}{2}mb_p = \sum_{r=1}^{m-1} y_r \sin prx_p$$

where $p = 1, 2, \ldots, n$ and m is odd.

But if m is even, the coefficient a_p (where $p = \frac{1}{2}m$) is

$$ma_{\frac{1}{2}m} = \sum_{r=0}^{m-1} y_r \cos r\Pi$$

Although these expressions for the coefficients of the periodic terms are adequate for the time series analysis, one more improvement can be made. Instead of limiting the range of the summations from 0 to (m-1), this range can be taken as the largest integral multiple of the period of the harmonic in the time series. The corresponding equations are given by Davis (3,57)

$$A(T) = \frac{2}{N'} \sum_{t=0}^{N'} y_t \cos \frac{2\Pi t}{T}$$

$$(1:4:7)$$

$$B(T) = \frac{2}{N'} \sum_{t=0}^{N'} y_t \sin \frac{2\Pi t}{T}$$

where T is the period of the harmonic considered

N' is chosen equal to the largest multiple of T in the time series.

1:5 The Case for "Detrending" the Time Series

In order to investigate the effect of the secular trend of the mean on the harmonic analysis consider the case of a linear trend. This development is given by Davis (3,75).

Suppose the time series in the interval

$$-a \leq t \leq a$$

has the trend

$$y = y_0 + mt$$

Suppose also that the harmonic analysis reveals that the time series has a harmonic term of the form

$$h(t) = A(T) \cos \frac{2\Pi t}{T} + B(T) \sin \frac{2\Pi t}{T}$$

where T is the period of the harmonic. If y, above, is expanded in a Fourier series in the interval

$$- a \leq t \leq a$$

the result is

$$y = y_0 + \frac{2ma}{\Pi} \{\sin \frac{\Pi t}{a} - \tfrac{1}{2} \sin \frac{2\Pi t}{a} + \frac{1}{3} \sin \frac{3\Pi t}{a} \ldots\} \qquad (1:5:1)$$

Now if in h(t) the period T belongs to the Fourier sequence, that is, if there is an integer n such that $n = 2a/T$, then the corresponding term in (1:5:1) must have been included in the coefficient of the sine term B(T) obtained by the Fourier analysis. Hence, the coefficient of $\sin(2\Pi t/T)$ which belongs to the true harmonic, independent of the trend, must be B(T) diminished by that part due to the trend.

Since the influence of the trend upon the harmonic is the term

$$(-1)^n \frac{2ma}{\Pi} \frac{1}{n} = (-1)^n \frac{mT}{\Pi}$$

the true harmonic is the function

$$h'(t) = A(T) \cos \frac{2\Pi t}{T} + B'(T) \sin \frac{2\Pi t}{T}$$

where

$$B'(T) = B(T) + (-1)^n \frac{mT}{\Pi} \qquad (1:5:2)$$

If σ^2 is the variance of the original series, then the variance σ_1^2 reduced by the trend and the harmonic term will be

$$\sigma_1^2 = \sigma^2 - \sigma_T^2 - \sigma_H^2$$

where σ_T^2 is the variance due to the trend

σ_H^2 is the variance due to the harmonic term

It has been shown in (1:4:4) that

$$\sigma_H^2 = \tfrac{1}{2}\left(A^2(T) + B'^2(T)\right) .$$

For the trend

$$\sigma_T^2 = \tfrac{1}{2} \frac{4m^2a^2}{\Pi^2} \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots\right) = \frac{m^2a^2}{3} .$$

If the series is defined over the interval

$$0 \leqq t \leqq 2a$$

instead of the interval

$$- a \leq t \leq a$$

then the only modification in the above analysis is merely that B'(T) as given in (1:5:2) is replaced by

$$B'(T) = B(T) + \frac{mT}{\Pi}$$

In the case in which 2a/T is not an integer, the period T does not belong to the Fourier sequence. In this case the above analysis will yield only an approximation to the reduced variance σ_1^2.

The above analysis for a linear trend can be easily extended to other types of trends. However, trend analysis described in section 1:2 provides a method for removing the secular trend of the mean from the time series. Hence, if the data is in the "detrended" form then harmonic analysis gives a true representation of the periodic component of the time series.

2.0 FORECASTING THE TIME SERIES

2.1 Moving Averages

Up to this point the object of the analysis has been to determine the forecasting model. The analysis has taken the following forms

(1) Trend analysis using regression to determine the trend of the mean and to put the data in "detrended" form.

(2) Autocorrelation analysis to determine the basic periocity of the time series.

(3) Spectral analysis to choose the periodic terms in the model

The next step in the analysis is forecasting the value of the time series for the future period. Considering again the representation of the time series

$$y(t) = \xi(t) + \varepsilon(t)$$

where $y(t)$ is the observed value of the time series at time t.

$\xi(t)$ is the process which the time series is following.

The criterion for the forecasting technique is to give an estimate of the process by effectively damping out the superimposed noise. Thus, a technique is needed which will seek a balance between the ability to respond to secular changes in the process and the effect of error in the forecast due to the random variation.

As an illustration of such a technique the moving average method can be considered. In this case the process is considered to be, at least locally, constant. Hence, the process can be described locally as

$$\xi(t) = a$$

By including the random noise the time series can be represented as

$$y_t = a + \varepsilon_t$$

The technique of the moving average gives the following estimate of the process (7,98)

$$M_t = \frac{y_t + y_{t+1} + y_{t+2} + \cdots + y_{t+N-1}}{N}$$

where M_t is the average of the N most recent periods in the data. If the autocorrelation analysis (1.3) shows that the time series is basically periodic then for the best results N should be some multiple of the period in order to negate the effect of periocity on the value M_t.

The rate of response of the moving average is controlled by the value of N. Since each of the N most recent observations is given the weight 1/N, as N is increased the response of the model to the most recent observation is decreased. This response can be seen more clearly by writing the recursive relationship for the moving average.

$$M_t = M_{t-1} + \frac{y_t - y_{t-N}}{N}$$

Suppose that the time series is following a constant process with superimposed noise about a mean a'. Then suddenly the process jumps to a new mean a". Brown notes that it would take N observations for the moving average to fully adapt to this change. (1,99)

2.2 Exponential Smoothing

The lag in the rate of response to a change in the process is one of the most critical characteristics of the forecasting technique. Moreover, even though the rate of response of the moving average can be altered by changing the value of N, the calculations involving this change must be carried out over the whole range of the observations. Thus a more palitable approach to the forecasting method is needed.

The recursive relationship for the moving average

$$M_t = M_{t-1} + \frac{y_t - y_{t-N}}{N}$$

can be approximated by the relation

$$\hat{M}_t = 1/N\ y_t + (1 - 1/N)\ M_{t-1}$$

where $\hat{M}_t$ is used to denote the estimated value of M_t. The underying assumption in this case is that y_{t-N} can be reasonably estimated by $1/N\ (M_{t-1})$. Brown uses S for smoothing instead of M for moving average and obtains the relation (1,107).

$$S_t(y) = \alpha y_t + (1-\alpha)S_{t-1}(y) \qquad (2.2.1)$$

where $S_t(y)$ is termed the smoothed statistic evaluated at time t.

α is an undimensioned ratio similiar, but not equal to 1/N.

The carrying out of the above relationship is called exponential smoothing.

By rearranging the expression for exponential smoothing an interesting observation can be made. Expressing (2.2.1) as

$$S_t(y) = S_{t-1}(y) + \alpha(y_t - S_{t-1}(y))$$

the current value of the smoothed statistic is expressed as the previous smoothed value plus a fraction of the difference between the value of the time series at the present time and the forecasted value at the previous period. This idea of updating the current estimate of the time series as a function of the error of the previous estimate is found to be quite consequential in later development.

Up to this point no sound justification is presented for the use of the smoothed statistic as a representation of the process. This justification is found in the definition of expectation. Hogg and Craig define expectation as follows (7). Let X be a random variable having a P.D.F. f(x), and let u(X) be a function of X such that

$$\int_{-\infty}^{\infty} u(x)f(x)dx$$

exists, if X is a continuous type of random variable, or such that

$$\sum_{x} u(x)f(x)$$

exists, if X is a discrete type of random variable. The integral, or the sum, as the case may be, is called the mathematical expectation (or expected value) of u(X) and is denoted by $E\left(u(X)\right)$. That is

$$E\left(u(X)\right) = \sum_{x} u(x)f(x)$$

if X is a discrete type of random variable

Brown uses expectation to obtain the following proof of the validity of exponential smoothing. (1,101) The following expansion is first performed

$$S(y) = \alpha y_t + (1-\alpha)\left(\alpha y_{t-1} + (1-\alpha)\ S_{t-2}(y)\right)$$

$$= \alpha y_t + \alpha(1-\alpha)\ y_{t-1} + (1-\alpha)^2\left(\alpha y_{t-2} + (1-\alpha)\ S_{t-3}(y)\right)$$

$$= \alpha y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \ldots$$

$$+ \alpha(1-\alpha)^n\ y_{t-n} + \ldots + (1-\alpha)^t\ y_0$$

$$= \alpha \sum_{k=0}^{t-1} (1-\alpha)^k\ y_{t-k} + (1-\alpha)^t y_0$$

The expected value of the above expression is then

$$E\left(S(y)\right) = \alpha \sum_0^{\infty} \beta^k E\left(y_{t-k}\right)$$

$$= E(y)\ \alpha \sum_0^{\infty} \beta^k = \frac{\alpha}{(1-\beta)}\ E(y) = E(y)$$

where, for convenience, $\beta=(1-\alpha)$. Hence, the justification of using the smoothed statistic as a forecast of the time series lies in the fact that the expected value of the smoothed statistic is the time series.

In the comparison of the technique of the moving average with exponential smoothing it is important to consider the weights given to the observations. In the case of the moving average the N most recent observations are given a weight of 1/N while all other observations are given weight zero. In

exponential smoothing the current observation is given a weight of α and the weight of all previous observations decreases geometrica-ly with age.

Moreover, in the moving average technique N observations must be carried "on the books" at all times. This can be a disadvantage when N is large and when a large number of time series are being considered. Furthermore, the moving average assigns no weight to any observation beyond the last N observations even though the contribution of these older terms may be significant. In contrast, exponential smoothing carries in one word of data all the history of the time series.

On the topic of sensitivity, it is a simple matter to change the value of α at any time and thus alter the response of the smoothed statistic. Again the moving average technique falls short. Although the response of the moving average can be altered by changing the value of N this change necessitates recomputation throughout the whole range of the data. On the basis of these comparisons further consideration of the moving average technique is ignored.

2.3 General Theorem of Exponential Smoothing

The previous analysis assumes that the process can be adequately represented by a constant model. The next step is the application of exponential smoothing to models other than the constant model. In the application of exponential smoothing to the constant model the recursive relationship (2.2.1)

$$\hat{a} = S_t(y) = \alpha y_t + (1-\alpha)\, S_{t-1}(y)$$

where the process is $\xi(t) = a$

gave a method of re-evaluating the estimate of the coefficient in the model with each observation. If this technique is extended to more complicated

models then a method must be determined for recursively re-evaluating the coefficients in those models. Brown develops the extension of the model to an n^{th} degree polynomial (1,132). In this case the process is represented by

$$\xi(t) = a_0 + a_1 t + a_2/2\ (t^2) + \ldots + a_n/n!\ (t^n)$$

The Taylor series expansion about the t^{th} observation yields an estimate of the future observations as follows

$$\hat{y}_{t+\tau} = \hat{y}_t^{(0)} + \tau \hat{y}_t^{(1)} + \frac{\tau^2}{2} \hat{y}^{(2)} + \ldots + \frac{\tau^n}{n!} y_t^{(n)}$$

where $y_t^{(k)}$ is the k^{th} derivative evaluated

at time t, (in this case t is taken to

be the current value)

$\hat{y}_t^{(k)}$ is an estimate of $y_t^{(k)}$

$$y_t^{(k)} = \left.\frac{d^k y}{dt^k}\right|_t$$

τ is the forecast interval

Thus the Taylor series expansion yields the following estimate for the next observation

$$y_{t+\tau} = \sum_{k=0}^{n} \frac{\tau^k y_t^{(k)}}{k!} = \sum_{k=0}^{n} \frac{\tau^k a_k}{k!}$$

Hence, the forecast is in terms of the current estimates of the derivatives of the model. These derivatives correspond to the coefficients that are required. The immediate goal then is to estimate these derivatives through the technique of exponential smoothing.

The expression for the smoothed statistic is given in (2.2.1) as

$$S_t(y) = \alpha y_t + (1-\alpha)\, S_{t-1}(y)$$

If this is referred to as single smoothing and double smoothing is defined as

$$S_t^{(2)}(y) = \alpha S_t^{(1)}(y) + (1-\alpha)\, S_{t-1}^{(2)}(y)$$

Then multiple smoothing of order k can be defined as

$$S_t^{(k)}(y) = S_t^{(k-1)}(y) + (1-\alpha)\, S_{t-1}^{(k)}(y) \qquad (2.3.1)$$

The fundamental theorem of exponential smoothing given in Appendix D states that if the observations y_t can be represented by the model

$$y_{t+\tau} = \sum_{k=0}^{n} \frac{\tau^k\, y_t^{(k)}}{k!}$$

then the general smoothed statistic can be represented as

$$s_t^{(P)}(y) = \sum_{k=0}^{n} (-1)^k \frac{y_t^k}{k!} \frac{\alpha p}{(p-1)!} \sum_{j=0}^{\infty} j^k \beta^j \frac{(p-1-j)!}{j!} \qquad (2.3.2)$$

The significance of the fundamental theorem to the goal of representing the time series by a polynomial is as follows.

(1) The general smoothed statistic was defined in (2.3.1).

(2) For any polynomial of degree n, by (2.3.2) n+1 smoothed statistics can be written in terms of the n+1 derivatives.

(3) Using the n+1 simultaneous equations from (2), the values of the derivatives $y_t^{(n)}$ can be solved for as linear combinations of the smoothed data.

Hence, a method is provided for recursively estimating the values of the coefficients in the polynomial model. Looking at the computational effort, however, for each observation in the series the n+1 smoothed statistics have to be recalculated and the n+1 simultaneous equations solved for the n+1 derivatives. The next effort, therefore, is to simplify these operations.

2.4 Matrix Representation of Exponential Smoothing

From Appendix D the fundamental theorem of exponential smoothing can be expressed in matrix form as

$$S_t = Ma \tag{2.4.1}$$

where S_t is the nx1 vector of smoothed statistics

a is the nx1 vector of coefficients

M is a nxp matrix with elements involving infinite sums of powers of the smoothing constant (where by virtue of the fundamental theorem n=p)

With the expression for the fundamental theorem expressed in the form (2.4.1) the vector of coefficients can be easily solved for as (1,137)

$$a = \hat{S}_t\ M^{-1}$$

where M^{-1} is the inverse of the corresponding square matrix

This type of recursive relationship is quite adaptive to computer programming.

Brown shows that just as in the case of the constant model the validity of the recursive relationship for the coefficients in the polynomial model can be proved by using expectation (1,138). Suppose that the time series can be described by the polynomial

$$y_t = a + bt + ct^2 + \dots + gt^n + \varepsilon_t$$

$$= \xi(t) + \varepsilon_t$$

where ε_t is the noise sample

$\xi(t)$ is the process

Since smoothing is a linear operation

$$S(y) = S(\xi) + S(\varepsilon_t)$$

But by definition

$$E\big(S(\varepsilon)\big) = 0$$

and it was shown that exponential smoothing yields the expected value of the data so that

$$E\big(S(y)\big) = S(\xi)$$

Up to this point the only forecasting model that has been developed is the polynomial. The reason for beginning with this limitation is, of course, that the fundamental theorem expresses the general smoothed statistic in terms of the derivatives

$$y_t^{(k)} = \frac{d^k y}{dt^k}\bigg|_t$$

which are the coefficients of the polynomial model. The polynomial model is by no means adequate in representing the time series. To be an effective representation of the process of the time series the model must contain terms which express both the trend component of the mean and the seasonal component. As seen in the section on representing the time series, the seasonal component is most adequately described by sine and cosine terms. Hence, a transition must be made for recursively estimating the coefficients in a more general model.

Suppose that the process can be represented by

$$\xi(t) = a_1 f_1(t) + a_2 f_2(t) + \ldots + a_n f_n(t)$$

$$= \sum_{i=1}^{n} a_i f_i(t) \qquad (2.4.2)$$

Where the functions $f_i(t)$ are of the types

(1) Polynomials

(2) Trigonometric functions

(3) Exponential functions

(4) Emperical functions

In some cases it might be advantageous to use fitting functions that are emperical such as the number of building contracts let 6 years ago. The only criterion that these emperical functions are required to meet is that their value be known both at the time the forecast is made and at the time in the future for which the forecast is required. However, emperical functions lead to computational difficulties far beyond those of the other types. Both because of these computational difficulties and for reasons of interest the emperical functions are avoided here.

The method for this general case is required to serve two objectives

(1) Provide a simple iterative proceedure for revising the estimates of the coefficients in the forccasting model

(2) Provide a means for discounting the weights given to the observations according to a time scale.

The expression for the forecast is given by Brown (1,161).

$$y(t+\tau) = \hat{a}_1(t)f_1(t+\tau) + \hat{a}_2(t)f_2(t+\tau) + \ldots$$

$$+ \hat{a}_n(t)f_n(t+\tau)$$

$$= \sum_{i=1}^{n} \hat{a}_i(t)f_i(t+\tau) \tag{2.4.3}$$

The residual in this case is defined as

$$y(T-j) - \hat{y}(T-j) = e(T-j) \tag{2.4.4}$$

where $y(T-j)$ stands for the model in which the coefficients are evaluated with all the data through time T but with the model evaluated j periods earlier

Appendix B gives the expression for the coefficients that minimize the sum

$$\sum_{t=11}^{T} w_t^2 \, e_t^2 \tag{2.4.5}$$

where w_t is the weight given the residual at time t as

$$a' = y \, W^2 \mathcal{Z}' \, F^{-1} \tag{2.4.6}$$

where W is a T x T matrix in which W_{ii} is the square root of the weight w_t given the residual for time i, and all off diagonal elcments of W are zero.

$\mathcal{Z}$ is an n x T matrix of elements $f_i(t)$, the value of the i^{th} fitting function at time t.

F is the n x n symmetric matrix

$$F = (\mathcal{Z}W)(\mathcal{Z}W)' = \sum_{t=1}^{T} w_t^2 \, f(t) \, f'(t)$$

If the data is discounted as in the case of exponential smoothing then the weight w in expressions (2.4.5) and (2.4.6) above must satisfy the relationship (1,163)

$$w_{T-j}^2 = \beta^j$$

and (2.4.5) becomes

$$\sum_{j=1}^{T} \beta^j \left(y(T-j) - \sum_{i=1}^{n} \hat{a}(T) \, f_i(T-j)\right)^2 \tag{2.4.7}$$

The F matrix becomes

$$F_{jk}(T) = \sum_{j=0}^{t-1} \beta^j \, f_i(T-j) \, f_k(T-j)$$

Hence a method for discounting the weights given to the observations and recursively updating the vector of coefficients is developed for the general model.

Since the major part of the calculations involved in updating the vector of coefficients is the formation of the F matrix, this matrix is chosen for further consideration. The value of the F matrix depends on three factors.

(1) The total number of observations in the time series, T.

(2) The fitting functions contained in the forecasting model

$$f_i(t) \quad i = 1, 2, \ldots, n$$

(3) The weighting function w_t^2 which in this case is given by the relationship

$$w_{T-j}^2 = \beta^j$$

Hence, the F matrix does not in any way depend on the values of the observations in the time series. Brown (1,163) uses this independence to develop a recursive relationship for the F matrix.

$$F(t) = f(t)f'(t) + \beta F(t-1) \qquad (2.4.8)$$

Referring to expression (2.4.6) the next computational effort to be considered is the formation of the data vector defined as

$$g(T) = \begin{bmatrix} g_1(T) \\ g_2(T) \\ \vdots \\ g_n(T) \end{bmatrix} = y\ W'\ W\ \mathcal{Z}$$

Then the i^{th} component of the data vector can be written as

$$g_i(T) = \sum_{j=0}^{T-1} \beta^j\ y\ (T-j)\ f_i(T-j) \qquad (2.4.8)$$

Brown (1,164) develops a recursive relation for this vector as

$$g_i(T) = y(T)\ f_i(T) + \beta\ g_i(T-1)$$

Hence, from (2.4.6), after n observations the coefficients can be estimated by

$$\hat{a}' = \hat{a}(T) = g(T)\ F^{-1}(T)$$

Hence, a reasonable forecast for the general model is (1,164)

$$y(T+\tau) = a'(T)f(T+\tau)$$

$$= \sum_{i=1}^{n} a_i(T)f_i(T+\tau)$$

Stopping for a moment to evaluate the progress of the forecasting development the following is noted.

(1) A scheme is developed for applying exponential smoothing to the case in which a locally constant process is assumed.

(2) The technique of exponential smoothing is extended from the case of a constant model to a general polynomial by means of the definition of general smoothing and the general theorem of exponential smoothing.

(3) Discounted multiple regression is introduced. This technique enables the further extension from the case of the general polynomial to a model which contains both polynomials and transendendental functions

2.5 Computational Considerations

At this point the development of the forecasting model is complete. That is, the time series under consideration can be adequately represented by a model composed of polynomials and transendentals. Next, consideration is given to improving the computational efficiency of the forecasting scheme.

If a comparison is made between the method of estimating the coefficients in the polynomial model and the method using discounted multiple regression an important difference is noted. In discounted multiple regression the coefficients of the model are estimated with respect to a fixed

time origin. On the other hand, in the case of the polynomial model time is measured with respect to the most recent observation.

Applying the concept of the moving time origin to discounted multiple regression and taking t+1 = T observations Brown (1,168) notes that the error criterion in (2.4.7) becomes

$$\min \sum_{j=0}^{t} \beta^j \left(f'(-j)a(t) - y(t-j)\right)^2$$

This is the same as the error criterion given in (2.4.7) except that time is counted with the current value as the origin.

If the same change in the time origin is applied to the data vector in expression (2.4.8) the result is

$$g_i(t) = \sum_{j=0}^{t} \beta^j f_i(-j)y(t-j) \qquad (2.4.8)$$

and in the same manner as the development for the fixed time origin the expression for the coefficients that minimize the error is

$$a' = y\ W^2 \mathcal{Z} F^{-1} \qquad (2.4.9)$$

with the criterion that there be at least n observations. The coefficients in the forecasting equation are estimated as before by

$$\hat{a}(t) = F^{-1}(t)g(t) \qquad (2.4.10)$$

2.6 Recursive Fitting Functions

With certain types of fitting functions the value of the vector of fitting functions can be obtained as linear combinations of the value of that vector at the previous time period. In the cases in which this recursive relationship holds the functions are said to have a fixed transition matrix.

That is, there are a set of coefficients L_{ij} which do not depend on time such that

$$f_1(t+1) = L_{11}f_1(t) + L_{12}f_2(t) + \ldots + L_{1n}f_n(t)$$

$$f_2(t+1) = L_{21}f_1(t) + L_{22}f_2(t) + \ldots + L_{2n}f_n(t)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$f_n(t+1) = L_{n1}f_1(t) + L_{n2}f_2(t) + \ldots + L_{nn}f_n(t) \qquad (2.6.1)$$

If the transition matrix is represented as L then (2.6.1) can be represented in matrix form as

$$f(t+1) = Lf(t) \qquad (2.6.2)$$

The only restriction placed on this transition matrix is that it have an inverse L^{-1}. The fitting functions for which such a transition matrix exists are the polynomials, exponentials and sinusoids. Hence, if the transition matrix is specified along with the vector of fitting functions at time t=0 then the value of the vector of fitting functions at any other time t can be determined by the relation

$$f(t) = L^t f(0) \qquad (2.6.3)$$

Three types of transition matricies used in combination are found to be quite useful for the time series considered. Brown (1,165) gives the transition matrix for a polynomial as an n x n matrix with ones on the diagonal, ones in the first element to the left of the diagonal, and zeros everywhere else. For example, the transition matrix and initial vector of fitting functions for the cubic model are

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad f(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Note: In the case of the polynomial model the coefficient is the binomial coefficient

$$\binom{t}{k} = \frac{t!}{(t-k)!\,k!}$$

If the fitting functions are trigonometric both the sine and cosine of each harmonic must be included (see Appendix A). Thus the fitting functions are

$$f_1(t) = \sin \omega t \qquad f_2(t) = \cos \omega t$$

Brown (1,166) gives the transition matrix and initial vector of fitting functions as

$$L = \begin{bmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{bmatrix} \qquad f(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The third type of transition matrix is for the case in which growing amplitudes and shifting phase angles are included in the periodic terms (see Appendix A). Suppose the example above is expanded to include the fitting functions

$$f_3(t) = t \sin \omega t \qquad f_4(t) = t \cos \omega t$$

Brown (1,166) gives the transition matrix and initial vector of fitting functions in this case as

$$L = \begin{bmatrix} \cos\omega & \sin\omega & 0 & 0 \\ -\sin\omega & \cos\omega & 0 & 0 \\ \cos\omega & \sin\omega & \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega & -\sin\omega & \cos\omega \end{bmatrix} \qquad f(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

The rule for generating transition matricies for forecasting models whose fitting functions are linear combinations of the three types of fitting functions described above is:

RULE FOR GENERATING GRAND TRANSITION MATRIX

Place the basic submatricies of the three types described above on the main diagonal of the grand transition matrix and fill in the required positions with zeros.

For example, suppose the model chosen to represent the time series is a growing sinusoidal model with a harmonic. The mathematical expression for the model is then

$$\xi(T+\tau) = (a_1 + a_2t) + (a_3 + a_5t)\sin(2\Pi t/12)$$
$$+ (a_4 + a_6t)\cos(2\Pi t/12) + a_7\sin(4\Pi t/12)$$
$$+ a_8\cos(4\Pi t/12)$$

where the basic period is 12. The basic submatricies, then, used in building the grand transition matrix must represent

(A) A polynomial with two degrees of freedom.

(B) A growing sinusoid with frequency

$$2\Pi t/12$$

(C) A harmonic sinusoid with frequency

$$4\Pi t/12$$

The schematic representation of the grand transition matrix is given in Fig. 2.1.

$$\begin{bmatrix} \begin{bmatrix} A \end{bmatrix} & & 0 \\ & \begin{bmatrix} \quad B \quad \end{bmatrix} & \\ 0 & & \begin{bmatrix} C \end{bmatrix} \end{bmatrix}$$

Fig. 2.1 Schematic of Grand Transition Matrix.

where the 0's represent the required zero level terms.

Making the appropriate substitutions the grand transition matrix is

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3}/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/2 & \sqrt{3}/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3}/2 & 1/2 & \sqrt{3}/2 & 1/2 & 0 & 0 \\ 0 & 0 & -1/2 & \sqrt{3}/2 & -1/2 & \sqrt{3}/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & \sqrt{3}/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{3}/2 & 1/2 \end{bmatrix}$$

Using the transition matrix and the new method of counting time the calculations for revising the estimates of the coefficients in the general forecasting model can be simplified. The expression for the data vector (2.4.8) can be written in the following form (1,169).

$$g(t) = y(t)f(0) + \sum_{j=1}^{t} \beta^{j} f(-j)\, y(t-j) \tag{2.6.4}$$

If successive values of the vector of fitting functions are generated with the transition matrix by the relation

$$f(-j) = L^{-1}f(-j+1) \tag{2.6.5}$$

expression (2.6.4) can be written as

$$g(t) = y(t)f(0) + \sum_{j=1}^{t} \beta^{j} L^{-1} f(-j+1)\, y(t-j) \tag{2.6.6}$$

By changing the index of summation by the relation k=j-1, the recursive relation for the data vector can be written as

$$g(t) = y(t)f(0) + \beta\, L^{-1} g(t-1) \tag{2.6.7}$$

In the above expression the effect of the new method of counting time can be seen. In expression (2.6.7) the current observation is weighted by the function vector f(0). In the previous method of counting time, with a fixed time origin, the current observation is weighted by the function vector f(t).

Two major improvements in the calculations are generated by changing the time origin and using the transition matrix.

(1) The only time dependent value in expression (2.6.7) is the observation. Hence, the components of the data vector no longer depend on absolute time and can be tabulated as constants.

(2) A simple recursive relationship for the data vector is developed.

Turning attention to the F matrix, Brown (1,170) determines the recursive relationship as

$$F(t) = \sum_{j=0}^{t} \beta^j f(-j)f'(-j) = F(t-1) + \beta^t f(-t)f'(-t) \tag{2.6.8}$$

Even with these simplifications, however, the computations have not yet significantly decreased. The major advantage in changing from a fixed time origin to a moving time origin is found in the following property of the F matrix. In the cases considered the fitting functions are either trigonometric functions or polynomials and β is less than one. Under these conditions β^t tends toward zero faster than $f(-t)$ can grow so that the F matrix reaches a steady state condition. Hence, F inverse can be determined in its final form for any set of fitting functions of the types considered. The expression used to describe this convergence is

$$F(t) = F(t-1)$$

Brown (1,170) notes the following properties of this convergence criterion. If a fitting function is used which takes the form of a decreasing exponential i.e.

$$f(t) = e^{-at},$$

the F matrix will reach a steady state only if past data is discounted at a very rapid rate, that is if

$$\beta < e^{-2a} .$$

Moreover, if the fastest growing function in the model is t^n, then the number of periods taken for this convergence is given approximately by

$$T = \frac{7 + (5.1)n}{(1-\beta)^{0.95}}$$

Since the steady state conditions are assumed to have been reached the time notation is dropped. Moreover, if the steady state condition is assumed the conditions for F to have an inverse will necessarily insue. Therefore the coefficients in the model can be estimated by

$$a(t) = F^{-1}g(t)$$

Therefore, the forecast of future observations is given by

$$\begin{aligned} y(t+\tau) &= a'(t)f(\tau) \\ &= \left(F^{-1}g(t)\right)'f(\tau) \\ &= g'(t)\ F^{-1}f(\tau) \\ &= g'(t)c(\tau) \end{aligned}$$

where τ is the forecast period

$g'(t)$ is the transpose of the current data vector

$c(\tau)$ is a column vector of coefficients that depend only on the values of the fitting functions at time τ, but not on absolute time

2.7 General Exponential Smoothing

The simplification of the calculations for the general model up to this point depend on the convergence of the matrix of weighted fitting

functions $F(t)$. This convergence expressed as

$$F = F(\infty)$$

requires two conditions

(1) Successive values of the vector of fitting functions can be generated by a fixed transition matrix L

(2) The origin of time is taken at the present

Furthermore, the data vector can be defined recursively by the relation (2.6.4)

$$g(t) = y(t)f(0) + \beta L^{-1} g(t-1)$$

Using these results the recursive estimates of the coefficients

$$\hat{a}(T) = (\hat{a}_1(T), \hat{a}_2(T), \ . \ . \ . \ , \hat{a}_n(T))$$

used in the forecast equation

$$\begin{aligned} y(T+\tau) &= a'(T)f(\tau) \\ &= \sum_{i=1}^{n} a_i(T)f_i(\tau) \end{aligned} \tag{2.7.1}$$

can be obtained.

From Appendix D the minimum discounted squared residual sum is attained when

$$F(T)a(T) = g(T)$$

Furthermore, when $F(T)$ has an inverse the vector of coefficients can be expressed as

$$a(T) = F^{-1}(T) \ g(T)$$

and if the convergence criterion is met, namely

$$f_i(t) < \beta^{-t/2} \quad \text{for all } i$$

the F matrix reaches a steady state

$$F = F(\infty) = \sum_{j=0}^{\infty} \beta^j \, f(-j)f'(-j)$$

Brown (1,177) substitutes this minimum steady state solution into the recursive relation for the data vector (2.6.4). The result is

$$Fa(T) = y(T)\, f(0) + \beta L^{-1}\, Fa(T-1)$$

If this expression is premultiplied by F^{-1}

$$a(T) = y(T)F^{-1}f(0) + \beta F^{-1}L^{-1}Fa(T-1) \tag{2.7.2}$$

This expression can be analyzed in the following manner. Defining the time independent vector

$$h = F^{-1}f(0)$$

and the time independent matrix

$$H = \beta F^{-1}L^{-1}F$$

expression (2.7.1) can be rewritten as

$$a(T) = hy(T) + Ha(T-1)$$

Considering

$$L^{-1}F$$

and postmultiplying the definition of the F matrix by $L'^{-1}L$

$$L^{-1}FL'^{-1}L' = \sum_{j=0}^{\infty} \beta^j \big(L^{-1}f(-j)\big)\big(L^{-1}f(-j)\big)'L'$$

$$= \frac{1}{\beta}\big(F-f(0)f'(0)\big)L'$$

Hence

$$H = \beta F^{-1}L^{-1}F = \big(I-F^{-1}f(0)f'(0)f'(0)\big)L'$$

But the h vector is defined as

$$h = F^{-1}f(0)$$

so

$$H = L' - h\big(Lf(0)\big)' = L' - hf'(1)$$

Moreover, (2.7.2) can be written as

$$a(T) = hy(T) + Ha(T - 1)$$

$$= hy(T) + L'a(T - 1) - hf'(1)a(T - 1)$$

Since $f'(1)\,a(T-1) = \hat{y}(T-1)$ is the forecast of what the observation at time T will be, as of the data received through time T-1. The above expression may be written as

$$a(T) = L'a(T - 1) + h\big(y(T) - \hat{y}(T - 1)\big) \tag{2.7.3}$$

and hence the final form for the recursive relation for the estimates of the coefficients in the general forecasting model is expressed as a function of the value of the vector of coefficients at the previous period and the error of the forecast made at the previous period.

Exponential smoothing provides an estimate of the coefficients $\hat{a}(T)$ from the observations $y(0), y(1), \ldots, y(T)$. If there is no noise in the data and if the fitting functions represent the process then if

exponential smoothing is applied over a long enough range the computed values of the coefficients will equal the true values. By definition, however, the time series does contain noise. The representation of the time series is given as

$$y(t) = \xi(t) + \varepsilon(t)$$

where $y(t)$ is the observed value at time t

$\xi(t)$ is the process

$\varepsilon(t)$ is the noise in the t^{th} observation.

The distribution of $\varepsilon(t)$ has the properties

$$E(\varepsilon_t) = 0 ,$$

$$E(\varepsilon_i \varepsilon_j) = 0 \quad \text{for } i \neq j$$

$$= \sigma^2 \text{ for } i = j,$$

where σ^2 is the variance of the noise distribution.

Exponential smoothing yields a forecast whose expected value is shown to be the process, $\xi(t)$. Brown (1,393) also uses expectation to investigate the estimates of the coefficients.

Suppose the process can be exactly represented by some linear combination of fitting functions

$$\xi = a$$

where the vector a is the true set of coefficients. The expected value of the forecast when the least squares criterion is used is shown to be

$$E(\hat{y}_t) = af(t)$$

Substituting the least squares estimates of the coefficients (2.4.9) into the expression for the expected value of the coefficients the following is

obtained.

$$E\ (a') = E(y\ W^2 \not{Z} F^{-1})$$

But since the expected value of the time series is the process

$$= \xi W^2 \not{Z} F^{-1}$$

$$= a' \not{Z} W^2 \not{Z} F^{-1}$$

$$= a'$$

Hence, the expected values of the estimates of the coefficients are the true coefficients.

The estimates of each coefficient, therefore, is some distribution whose mean is the true value of that coefficient. Next, the variance of the distribution is considered. The correlation coefficient between two random variables X_1 and X_2 as defined previously is

$$\rho_{12} = \frac{E\ (X_1 - \mu_1)\ (X_2 - \mu_2)}{\sigma_1 \sigma_2}$$

where μ_1 is the expected value of X_1

μ_2 is the expected value of X_2

σ_1 is the variance of X_1

σ_2 is the variance of X_2

Hogg and Craig (7) define the covariance of these variables as

$$E\left((X_1 - \mu_1)\ (X_2 - \mu_2)\right)$$

Moreover, the variance - covariance matrix is defined by Hogg and

Craig (7) as follows. Given the N random variables

$$X_1, X_2, \ldots, X_N$$

and calling the variance - covariance matrix V, the elements on the principal diagonal of V are respectively the variances $\sigma_{ii} = \sigma_i^2$, $i = 1, 2, \ldots, N$. The elements not on the principal diagonal of V are the covariances

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

Exploiting the idea of the variance - covariance matrix, the corresponding matrix for the variation in the estimated coefficients $\hat{a}$ caused by the noise $\varepsilon_t = y_t - \xi_t$ in the observed data is

$$E(\hat{a} - \tilde{a})(\hat{a} - \tilde{a})' = F^{-1} \; W'^{2} \; (x - \xi)'(x - \xi) \; W^2 \mathcal{Z}' \; F^{-1}$$

Since the noise is defined to have no serial correlation and assuming that all the noise samples have the same variance σ_ε^2, then (1,393)

$$E(x - \xi)' \; (x - \xi) = I\sigma_\varepsilon^2$$

and the covariance matrix for the coefficients reduces to

$$F^{-1}\mathcal{Z}W^2 \; (\mathcal{Z}W^2)' \; F^{-1}\sigma_\varepsilon^2 = F^{-1} \; K \; F^{-1} \; \sigma_\varepsilon^2 \qquad (2.7.4)$$

$$\text{where } K = (\mathcal{Z}W^2) \; (\mathcal{Z}W^2)'$$

The K matrix, then, can be represented as

$$K = \sum_{j=0}^{\infty} \beta^{2j} \; f(-j) \; f'(-j) \qquad (2.7.5)$$

Moreover, the variance - covariance matrix for the coefficients can be expressed in terms of the variance of the noise as

$$V\sigma_\varepsilon^2 = F^{-1} K F^{-1}\sigma_\varepsilon^2 \tag{2.7.6}$$

The (i,j) element of V_{ij} is the covariance between a_i and a_j

$$\text{cov}\{a_i, a_j\} = V_{ij}\sigma_\varepsilon^2$$

where σ_ε^2 is the variance of the (uncorrelated) noise. The variance of the i^{th} coefficient is

$$\text{var}\{a_i\} = \text{cov}\{\hat{a}_i, \hat{a}_i\} = V_{ii}\sigma_\varepsilon^2$$

Hence, the elements V_{ii} are the variances of the coefficients expressed as a multiple of the variance of the noise σ_ε^2.

3.0 COMPUTER PROGRAMS

The analysis presented in the first two chapters, trend analysis, autocorrelation analysis, spectral analysis and exponential smoothing can only be carried out practicably through the use of automatic computing equipment. The facility available for this study is the I.B.M. 1410 system at Kansas State University. This system consists of the following equipment; an I.B.M. 1410 computer with 40K storage capacity, a 1423 card reader and card punch, a 1422 printer, an I.B.M. 1401 computer and seven 7330 magnetic tape drive units.

The 1410 system is internally programmed with PR-155. This system allows programming in either Autocoder or Fortran. In this study Fortran is used. The processor occupies 10K leaving 30K for the compiled program and the calculations. This limitation of storage would be prohibitive to the application of the preceeding analysis except that the PR-155 system allows phasing of the program. Phasing consists of writing the program, which itself is too large for the available memory capacity, into parts, or phases. These phases are then run independently and anything that must be retained from one phase for following phases is read onto a "scratch" file. At the completion of a phase the processor automatically clears core and loads in the next phase.

Phasing the program does, however, have its associated limitations. There are, basically, two limitations which must be considered. The first limitation of phasing is concerned with time. Since each phase is compiled independently, compiling time is increased. Moreover, running time is increased because the data which must be retained between phases must be written on and read off tapes. The second, and most crucial limitation is concerned

with the nature of the program. Phasing is limited to programs which can be devided into independent parts which can be accomodated in the memory capacity of the available facilities.

The programs incorporated in this analysis provide a twofold function. First, these programs serve the function of a medium for carrying out and evaluating the analysis of the first two chapters. Secondly, these programs serve the function of analyzing the practicability and economy of applying the forecasting techniques to a digital computing system with a Fortran type of processor.

3.1 The program for general exponential smoothing

In the application of general exponential smoothing to computer simulation a general program is required. This program must have the ability to (1) perform general exponential smoothing on a time series, (2) change the forecasting model and (3) vary the significant parameters in the forecasting model. In the last requirement these parameters are taken to be, (a) the basic period of the model and (b) the value of the smoothing constant. Moreover, it is required that the program for the application of general exponential smoothing perform these functions in a reasonable amount of time, with the ability to handle a wide range of fitting functions and time series and provide output in the desired form.

Considering the operations involved in general exponential smoothing the following independent phases are suggested. The first phase, phase I, is named INCONT and executes the functions of

(1) Reading the time series into memory,

(2) Providing the description of each forecasting model and

(3) Evaluating estimates for the initial values of the coefficients for cach model and time series combination.

Upon completion of the first phase the following information will be available to the insuing phases

(1) Control parameters

(a) The basic period of the forecasting model, P

(b) The number of fitting functions, N

(c) The number of observations in the time series, ND

(2) The time series X(I), where I = 1,2,...,ND

(3) The transition matrix TM(I,J)

where I=1,2,...,N

J=1,2,...,N

(4) The initial vector of fitting functions, F(I), where I=1,2,...,N

(5) The initial value of the vector of coefficients C(I), where I =1,2,...,N

(6) The change vector, CHK(I), I=1,2,...,N.

Evaluating the above requirements 1,2 and 4 can be adequately performed by normal read operations. If a general program is to be maintained, however, requirement 3 and 5 must be considered more thoroughly. Considering requirement 3, it was noted in Chapter 2 that the elements of the transition matrix which are included to describe the periodic terms in the model take the form

$$\pm \cos \omega \ , \ \pm \sin \omega$$

where $\omega = 2\Pi/P$, and P is the basic period of the model. Hence, these terms must be adaptable to a change in the basic period of the model. Moreover, the initial values of the coefficients for the periodic terms in the forecasting model are given by Brown (1,194) as

$$a_n = 2/P \sum_{k=1}^{P} y_k \sin \omega_k$$

for the coefficient of the sine terms and

$$a_n = 2/P \sum_{k=1}^{P} y_k \cos \omega_k$$

for the coefficients of the cosine terms. Hence, provision for calculating 3 and 5 for each forecast is made.

Considering 6, the change vector has not yet been discussed. In section 2.6 the recursive relation for the vector of fitting functions is given in (2.6.2) as

$$f(t+1) = Lf(t)$$

where L is the transition matrix and f(t) is the vector of fitting functions evaluated at time t. However, because of the moving time origin the F matrix for general exponential smoothing is defined in (2.6.4) as

$$F(t) = \sum_{j=0}^{t} \beta^j f(-j)f'(-j) = F(t-1) + \beta^j f(-t)f'(-t)$$

The transition between the recursive relation for the vector of fitting functions defined for the fixed time origin and the recursive relation for the same vector defined for a moving time origin results in expression (2.6.5)

$$f(-t) = L^{-1}f(-t+1)$$

Hence, it would seem that the requirements of the F matrix necessitate storing the inverse of the L matrix. The objection with using the inverse of the transition matrix is based on the recursive relation for the vector of coefficients (2.7.3)

$$a(T) = L'a(T-1) + h\left(y(T) - \hat{y}(T-1)\right).$$

This relationship is executed in the last phase of the program. Hence, the original transition matrix and its inverse would have to be carried in the program. This dual storage is an obvious burden, if not a limitation, on the program.

Thus, the change vector is introduced to avoid the calculation and storage of the inverse of the transition matrix. Advantage is taken of the trigonometric identities

$$\cos(-x) = \cos(x)$$

$$\sin(-x) = -\sin(x)$$

and the relation

$$(-t)^n = \begin{cases} t^n \text{ for n even} \\ -(t^n) \text{ for n odd} \end{cases}$$

The fitting functions used for the time series considered are either simple powers of t or trigonometric functions or multiples of these functions. In the case of the general polynomial

$$y = a_0 + a_1(t) + a_2(t^2) + a_3(t^3) + \ldots + a_n(t^n)$$

the vector of fitting functions is

$$\begin{bmatrix} (1) \\ (-t) \\ (-t^2) \\ (-t^3) \\ \cdot \\ \cdot \\ \cdot \\ (-t^n) \end{bmatrix}$$

As an illustration of a more complicated model the example in section 2.6 can be considered. The expression for this model is given as

$$\xi(T+\tau) = (a_1 + a_2 t) + (a_3 + a_5 t)\sin(2\Pi t/12)$$

$$+ (a_4 + a_6 t)\cos(2\Pi t/12) + a_7 \sin(4\Pi t/12)$$

$$+ a_8 \cos(4\Pi t/12)$$

The vector of fitting functions in this case is

$$\begin{bmatrix} 1 \\ (-t) \\ \sin(\omega_1(-t)) \\ \cos(\omega_1(-t)) \\ (-t)\sin(\omega_1(-t)) \\ (-t)\cos(\omega_1(-t)) \\ \sin(\omega_2(-t)) \\ \cos(\omega_2(-t)) \end{bmatrix}$$

where $\omega_1 = 2\Pi t/12$ and $\omega_2 = 4\Pi t/12$. From these examples it can be seen that

$$f_i(-t) = \pm f_i(t)$$

where $f_i(t)$ is the value of the i^{th} fitting function evaluated at time t. In order to facilitate this relationship in the computer program the change vector CHK(I), where I = 1,2,...,N, is introduced. This vector satisfies the relationship

$$\mathrm{CHK(I)} \begin{cases} = 0 & \text{for } f_i(-t) = f_i(t) \\ = 1 & \text{for } f_i(-t) = -f_i(t) \end{cases}$$

Thus, the transition from a fixed time origin to a moving time origin is made without the use of the inverse of the transition matrix.

The functional value of this vector can be seen in the following example. Suppose the transition matrix is of the order 10 x 10. If a percision of ten decimal places is used this matrix, or its inverse, would require

$$10 \times 10 \times 10 = 1000$$

core locations for storage. In contrast the storage of the change vector of fixed point numbers requires only 10 core locations.

Finally, all the required information from the first phase is written onto a work tape, making it available for following phases. The flow diagram for phase I is shown in Fig. 3.1 and the program is shown in Appendix E.

The second phase, named RAY, has three functions, they are

(1) Calculating the F matrix,

(2) Calculating the K matrix, and

(3) Checking for convergence of the F matrix.

The expression for the F matrix as given in (2.6.8) is

$$F(t) = \sum_{j=0}^{t} \beta^j f(-j)f'(-j) = F(t-1) + \beta^t f(-t)f'(-t)$$

As shown in section 2.6 this matrix converges under specified conditions. That is, $F = F(\infty)$, or $F(t) = F(t-1)$. The convergence criteria used in this application is that suggested by Brown (1,)

$$\frac{F_{ij}(t) = F_{ij}(t-1)}{F_{ij}(t-1)} \leq 10^{-6}$$

for all i and j. Moreover, the K matrix is given in (2.7.5) as

$$K = \sum_{j=0}^{\infty} \beta^{2j} f(-j) f'(-j)$$

Again a recursive relation for this matrix can be formulated as

$$K(t) = K(t-1) + \beta^{2t} f(-t) f'(-t) \ .$$

The calculations required to form these matricies are quite extensive. Therefore, it is necessary to reduce the calculation time for these operations. The first reduction in computing time can be attained by noting that both the F and K matricies are symmetric. Moreover, the recursive relation for the K matrix can be expressed as

$$K(t) = K(t-1) + \left(\beta^{t} f(-t) f'(-t)\right) \beta^{t}$$

By defining the term

$$\beta^{t} f(-t) f'(-t) = Z(t)$$

the two recursive relationships can be written as

$$F(t) = f(t-1) + Z(t)$$

$$K(t) = K(t-1) + \beta^{t} Z(t)$$

But since Z is also a symmetric matrix the calculations are reduced to computing half of Z for each iteration. Moreover, by taking advantage of the recursive relationship

$$\beta^{t} = (\beta^{(t-1)})\beta$$

the calculations are further reduced.

The check for convergence requires the formation of the quotients

$$Z_{ij}(t)/F_{ij}(t-1)$$

for $i=1,2,\ldots,n$ $j=1,2,\ldots,n$, and checking to see if these terms exceed 10^{-6}. The calculation for this convergence check can be reduced in two ways

(1) By noting that the convergence criteria must be met for all the elements in the F matrix, the program can be written to exit from this check routine as soon as it finds one element which has not converged.

(2) The operations involved in checking for convergence are quite extensive. If this check is made each time the F matrix is updated the calculation time will be increased considerably. Hence, the program can be written to make this check at specified intervals. These intervals are taken to be each 50^{th} iteration.

The flow diagram for phase II is given in Fig. 3.2 and 3.3 and the program is given in Appendix E.

The third phase, named MATINV, performs the functions of

(1) Taking the inverse of the F matrix,

(2) Forming the h vector and

(3) Determining the variance of the coefficients.

The recursive relation for the coefficients using general exponential smoothing is given in equation (2.7.3) as

$$a(T) = L'a(T-1) + h\{y(T) - \hat{y}(T-1)\}$$

where

$a(T)$ is the estimate of the coefficients at time T ,

L is the transition matrix,

h is the h vector,

$\{y(T) - \hat{y}(T-1)\}$ is the error of the forecast made at the previous period.

The vector h is defined as

$$h = F^{-1}f(0),$$

where F^{-1} is the inverse of the F matrix and $f(0)$ is the vector of fitting functions evaluated at time $t=0$. It is found that the most convenient method for performing the three functions required by this phase is to write the program for determining the h vector as a subprogram of the main program for finding the inverse of the F matrix and calculating the variance of the coefficients. Again, taking consideration of the memory requirements it is noted that in future operations the F matrix is not needed. That is, only the inverse of the F matrix will be required after this phase. Hence, space can be conserved in core by replacing the F matrix by its inverse. This is equivalent to reading the inverse of the F matrix over the F matrix.

In section 2.7 the variance-covariance matrix is defined as

$$V\sigma_\varepsilon^2 = F^{-1}KF^{-1}\sigma_\varepsilon^2$$

where σ_ε^2 is the variance of the noise distribution. By using this matrix the variance of the coefficients expressed as multiples of the variance of the noise can be obtained from

$$\text{var}\{a_i\} = \text{cov}\{a_i, \hat{a}_i\} = V_{ii}\sigma_\varepsilon^2$$

That is, the elements on the diagonal of the covariance matrix provide the required information. It is found, however, more convenient to calculate the entire V matrix than to calculate the diagonal elements alone. The flow

diagram for phase III and the subprogram for calculating the h vector are shown in Fig. 3.4 and 3.5, and the programs are shown in Appendix E.

The fourth phase is the phase which actually makes the forecast. With the h vector calculated all the information required for this phase is available. The initial estimates of the coefficients are obtained in phase I. With these starting values the format for forecasting the time series is

(1) Make the forecast according to (2.7.1)

$$y(T+\tau) = a'(T)f(\tau)$$

$$= \sum_{i=1}^{n} a_i(T)f_i(\tau)$$

where $y(T+\tau)$ is the forecast for one period in the future,

$a(T)$ is the estimate of the vector of coefficients made at the present period,

$f(\tau)$ is the vector of coefficients evaluated at time τ - where τ is the forecast period.

(2) Update the vector of coefficients in terms of

(a) The previous vector of coefficients and the forecast period.

(b) The error of the forecast made in the previous period.

The recursive relation for the vector of coefficients is given in equation (2.7.2)

$$a(T) = L'a(T-1) + h\left(y(T) - \hat{y}(T-1)\right)$$

where $\hat{a}(T)$ is the present estimate of the vector of coefficients,

$a(T-1)$ is the vector of coefficients used to make the forecast in the previous period,

L' is the transpose of the transition matrix,

h is the vector of constants defined in section 2.7,

$y(T)$ is the present value of the time series,

$\hat{y}(T-1)$ is the forecast made at the previous period.

(3) Return to step (1)

The two auxiliary functions of the fourth stage are to

(1) Calculate the sum of squares of errors and

(2) Calculate the variance of the forecasts.

The expression for the variance of the forecasts is given in section 4.4 as

$$\sigma_F^2 = \frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{N} y_t}$$

where y_t is the observation at time t and $\hat{y}_t$ is the forecast made for time t. The flow diagram for phase four is given in Fig. 3.6 and the Fortran program is given in Appendix E

The last phase of the program, phase V, makes a plot of

(1) The time series,

(2) The forecasts and

(3) The absolute error of the forecast.

This phase is named "PLOTTER". The actual plotting is done in a subroutine "PLOTS".

It is found that a serious limitation is imposed on the length of the time series that can be accomodated in the program if the value of the forecast, the observations and the errors are stored in memory at one time. In the preceeding phase the forecast and the error of the forecast were calculated for each period. However, if these sets of values were stored in this phase the length of the time series that could be used would be too short for the investigation. Hence, some method must be devised to make the values of the forecasts and the forecast errors available to the plotting phase without storing them in the previous phase. In the case of the forecast this problem was solved in the following manner. Referring to Fig. 3.6, as each value of the forecast is calculated in phase IV it is written onto a work tape. There will be then, at the completion of this phase, N forecasts on the tape for each model used; where N is the number of observations in the time series. In the plotting phase these values can be read in and transformed into the subscript variable FCST(I), where I=1,2,...,N, making them available for the subroutine "PLOTS". In the case of the errors this difficulty is overcome since the subroutine "PLOTS" calculates the absolute value of the errors independently along with plotting the three values.

This subprogram provides a vertical plot of the three variables. The horizontal axis on which the three variables are measured is limited by the 1422 printer. The printer is capable of printing 133 characters on a line. Hence, the scale for the three variables must be transformed to an integer scale with a range of from zero to 133. In this case an upper limit of 130 was actually used. The vertical scale contains one line for each period

in the time series. This scale does not have an upper limit.

The plotting function is carried out in Fortran IV by means of a write statement. For each period in the time series the statement

```
WRITE(3,1) (MP(L),(L=1,130))
```

is executed. This statement can be inturpreted as writing, by means of the 1422 printer (symbolic unit 3), by the accomodating format (7), the subscripted variable MP(L) which ranges from 1 to 130. Now if for each line of the plot L takes on only three values corresponding to the observation, the forecast and the error for that period on the integer scale, then the required plots can be obtained. One of the characteristics of Fortran IV is that a variable can be set equal to an alphebetic or special character. Hence, if a method is provided to set each of the three values of the subscripted variable MP(L) equal to the corresponding symbol for the plot, then the requirements for a plotting routine are fulfilled. The flow diagrams for the program "PLOTTER" and the subprogram "PLOTS" are shown in Fig. 3.7 and 3.8 and the programs are given in Appendix E.

The second phase of the program is the most critical with respect to time. The controlling factors in this phase are the value of the smoothing constant and the size of the matrix being handled. The calculation time increases factorialy as the order of the matrix.

The relationship of the value of the smoothing constant to calculation time can be demonstrated as follows. The recursive relation for the elements of the F matrix is given in equation (2.6.5) as

$$F(t) = F(t-1) + \beta^{t} f(-t) f'(-t)$$

Since for any given set of fitting functions the rate of growth of $f_i(-t)$ is fixed the rate of convergence of the F matrix depends on the speed at which β^t goes to zero. From the experience of this investigation a value of the smoothing constant of 0.7 will yield convergence for the F matrix for a given set of fitting functions five times as fast as a value of 0.9.

In this section the flow diagrams for the separate phases of the forecasting program are shown. It is noted that phasing the program depends on the ability of the system to retain information between the phases by means of "work" or "scratch" tapes. Moreover, it is noted that phasing also depends on the ability to write the program in independent parts or phases that can be run separately. In this investigation one more limitation on phasing a program is found. Phasing the program depends on the ability to write the program into phases in such a way that the information from any phase can be made available for the following phase or phases which require it. This limitation can best be illustrated in the context of the program under investigation. Due to the complexity of the internal read and write statements between phases no attempt was made to represent them in the previous flow diagrams of this section. Figure 3.9 shows the data transfer statements for the "scratch" files. In this program three "scratch" files are required. These tapes are referred to in the listings of the forecasting program in Appendix E as symbolic units 5,6 and 7. It is found that it would not be possible to use any less than three tapes since the available memory capacity of the system would not allow the required data transfer.

The read and write statements shown in Fig. 3.9 are given in the program using free style formats. Thus each of the read and write statements

refer to a physical record representing one model. This technique is necessary to permit a multimodel program. This type of multiphase programming is considered to be a significant contribution of this investigation.

As a check on the validity and accuracy of the results obtained from the program for general exponential smoothing Table 3.10 is constructed. This table gives the values of the h vector and the variance of the coefficients for several models. For each model these values are obtained using three values of the smoothing constant 0.70, 0.90 and 0.95. The results are shown to 6 decimal places and the difference between those obtained by Brown (1,184-193) are shown. This table indicates that the results obtained are in essential agreement with those obtained by Brown.

LINEAR MODEL

$1 - \beta^n = 0.25$

h vector	diff
.250000	.000000
.017949	.000000
Variance of Coefficients	**diff**
.169367	.000000
.007400	.000000

$1 - \beta^n = 0.90$

h vector	diff
.100001	.000000
.002633	.000000
Variance of Coefficients	**diff**
.064458	.000000
.000036	.000000

$1 - \beta = 0.05$

h vector	diff
.0500013	.000000
.000641	.000000
Variance of Coefficients	**diff**
.031729	.000000
.000004	.000000

Table 3.10 Comparison of Results Obtained from Computer Program with Brown's Results

LINEAR MODEL WITH SUPERIMPOSED SINUSOID

$1 - \beta = 0.25$		$1 - \beta^n = 0.10$	
h vector	diff	h vector	diff
.1129498	.000000	.050242	.000000
.004568	.000000	.000659	.000001
.041148	.000001	.006054	.000000
.120502	.000018	.049758	.000012
Variance of Coefficients	diff	Variance Of Coefficients	diff
.085101	.000000	.032447	.000000
.000086	.000000	.000004	.000000
.0068774	.000016	.026124	.000016
.070726	.000024	.026316	.000017

Table 3.10 continued

3.2 The Program for Calculating the Autocorrelation Function and "Detrending"

In section 1.3 the autocovariance is defined as

$$R_{xx}(k) = \sum_{j=k+1}^{T} y_j y_{j-k} \;/\; (T-k+1)$$

where k is the lag for which the autocovariance is calculated

T is the range of the series

y_j is the series which has been adjusted to have an expected value of zero

The normalized form of the autocovariance is defined as

$$\rho(k) = R_{xx}(k)/R_{xx}(0)\ .$$

This is the autocorrelation coefficient. The set of values for the autocorrelation coefficient for all lags, k = ± 1, ± 2, ... , is defined as the autocorrelation function.

This expression as it stands does not readily lend itself to computer programming. A method for adapting the calculation of the autocorrelation function to computer programming is developed by Raymond W. Southworth (8). This method is based on the definition of the autocorrelation coefficient as

$$\rho(k) = \frac{(N-k)\sum_{i=1}^{N-k} y_i y_{i+k} - \left(\sum_{i=1}^{N-k} y_i\right)\left(\sum_{i=1}^{N-k} y_{i+k}\right)}{\left((N-k)\sum_{i=1}^{N-k}(y_i)^2 - \left(\sum_{i=1}^{N-k} y_i\right)^2\right)^{1/2}} \times \left((N-k)\sum_{i=1}^{N-k}(y_{i+k}^2) - \left(\sum_{i=1}^{N-k} y_{i+k}\right)^2\right)^{1/2} \qquad (3.2.1)$$

If the data is first put in the normalized form then this expression reduces to the one above.

The value of expression (3.2.1) is found in its adaption to computer programming. First the following sums are defined.

$$T_k = \sum_{i=1}^{N-k} y_{i+k} \qquad G_k = \sum_{i=1}^{N-k} y_i^2$$

$$F_k = \sum_{i=1}^{N-k} y_i \qquad C_k = \sum_{i=1}^{N-1} y_i y_{i+k}$$

$$S_k = \sum_{i=1}^{N-k} y_{i+k}^2 \qquad W(k) = C_k/N-k$$

Then it is noted that recursive relationships can be developed for the first four sums.

$$T_k = T_{k-1} - y_k$$

$$F_k = F_{k-1} - y_{N-k+1}$$

$$S_k = S_{k-1} - y_k^2$$

$$G_k = G_{k-1} - y_{N-k+1}^2$$

The flow diagram for the computer program for calculating the autocorrelation function is given in Fig. 3.11. This flow diagram illustrates how these recursive relationships reduce the calculations extensively.

This program also removes the trend of the mean from the time series. The "detrending" is performed in a subprogram named "TREND". As described in section 1.2 this operation basically consists of fitting a polynomial regression model chosen to represent the trend of the mean, to the data.

This regression curve is then subtracted from the data. In the time series considered in this investigation the linear regression model is found to be adequate. This model can be represented as

$$y(t) = a + b(t)$$

where y(t) is the value of the regression model at time t and a and b are the coefficients to be estimated by regression

The expression for the constant term in this model is given in Appendix B as

$$a = \sum_{i=1}^{N} y(i)/N$$

where N is the number of observations in the time series. The expression for the coefficient for the linear term is given as

$$b = \frac{\sum_{i=1}^{N} (i)y_i - (\bar{i}) \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} (i)^2 - 2(\bar{i}) \sum_{i=1}^{N} (i) + \sum_{i=1}^{N} (\bar{i})^2}$$

Since the independent variable in this case is the uniform time axis this expression can be reduced to

$$b = \frac{\sum_{i=1}^{N} (i)y_i - 1/N \sum_{i=1}^{N} (i) \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} (i)^2 - 1/N(\sum_{i=1}^{N} (i))^2}$$

The flow diagrams for the program to calculate the autocorrelation function and the subprogram to "detrend" the time series are given in Fig. 3.11 and 3.12 respectively. The programs are given in Appendix E.

3.3 The Program for Calculating the Power Spectrum

The basic functions of this program are calculating the coefficients of the harmonics of the time series and calculating the energy term for each frequency in the spectrum. The expressions for the coefficients are given in section 1.4 as

$$A(T) = 2/N' \sum_{t=0}^{N'} y_t \cos(2\Pi t/T)$$

$$B(T) = 2/N' \sum_{t=0}^{N'} y_t \sin(2\Pi t/T)$$

where T is the period of the harmonic,

N' is the largest multiple of T in the series.

In section 1.4 the energy term for measuring the contribution of the frequencies is given as

$$E(T) = \frac{R^2(T)}{2\sigma^2}$$

where $R^2(T) = A^2(T) + B^2(T)$,

T = the period of the harmonic,

σ^2 = the variance of the data.

For convenience the value of R(T) is taken as a measure of the contribution of the frequencies. This is compatable with the spectral analysis found in the literature.

The input to this program is the detrended data from the program in section 3.2. The flow diagram for the program to calculate the power spectrum is given in Fig. 3.13 and the program is given in Appendix E.

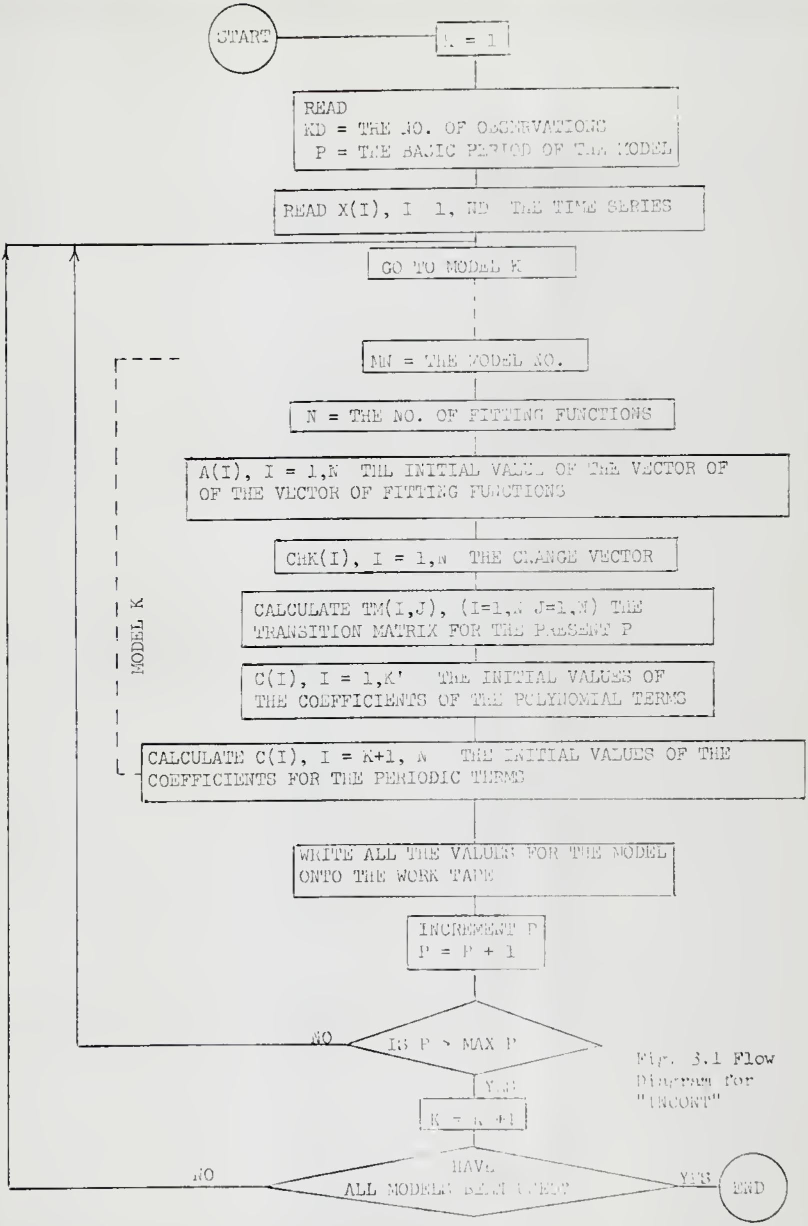

Fig. 3.1 Flow Diagram for "INCONT"

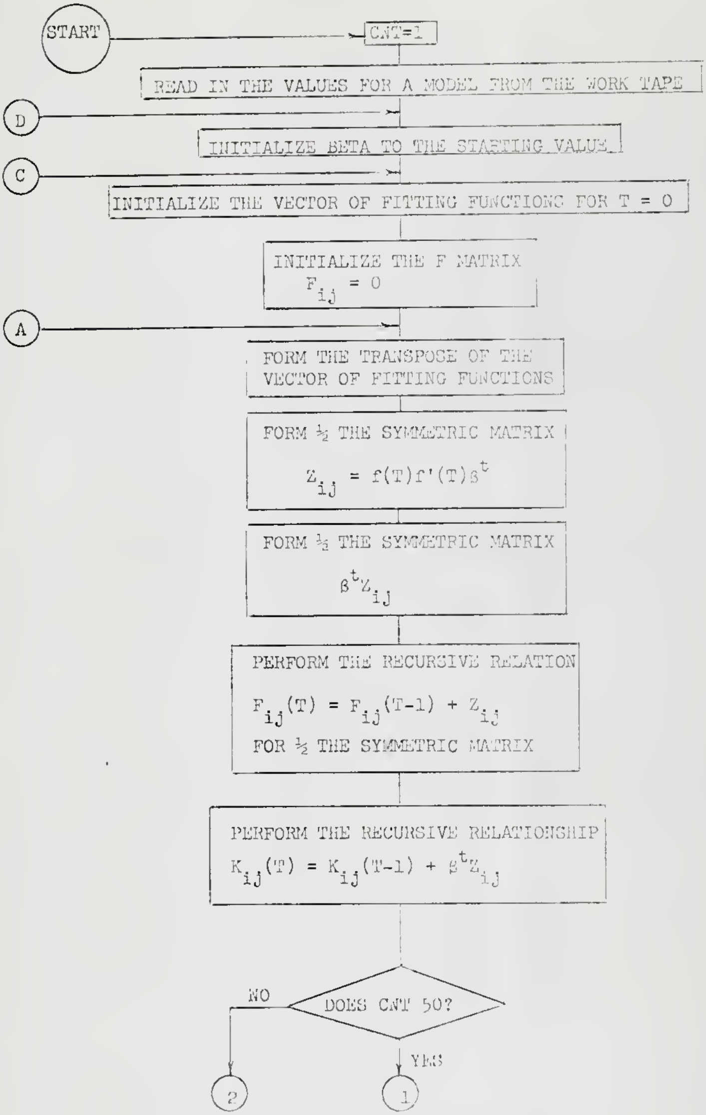

Fig. 3.2 Flow Diagram for Phase "RAY"

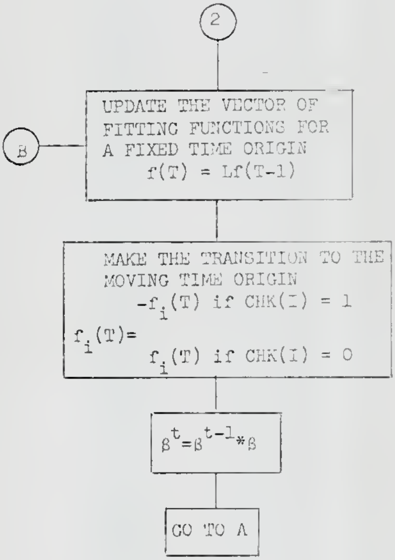

Fig. 3.2 continued

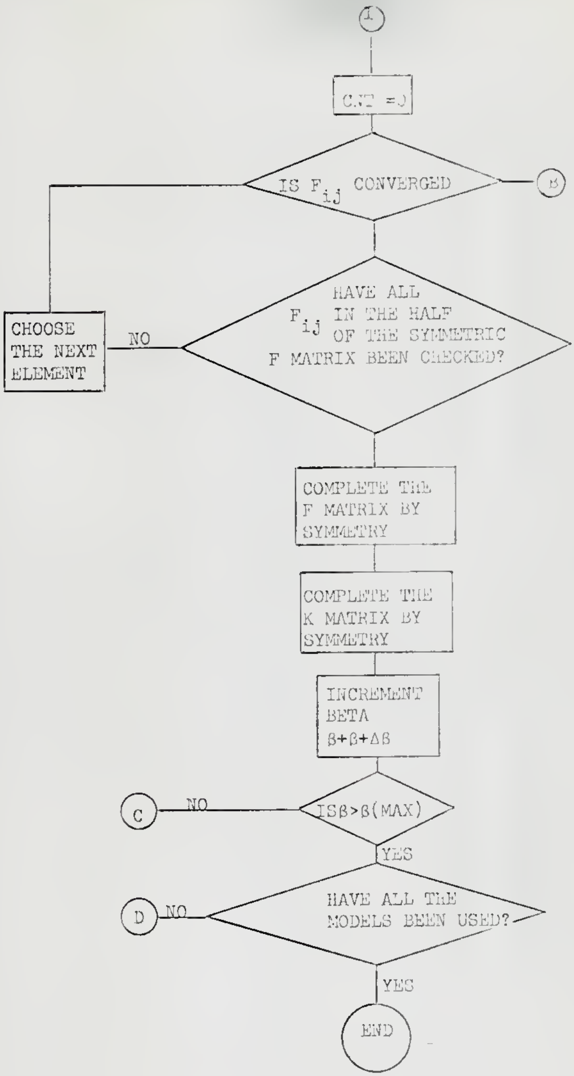

Fig. 3.3 Convergence Check for "RAY"

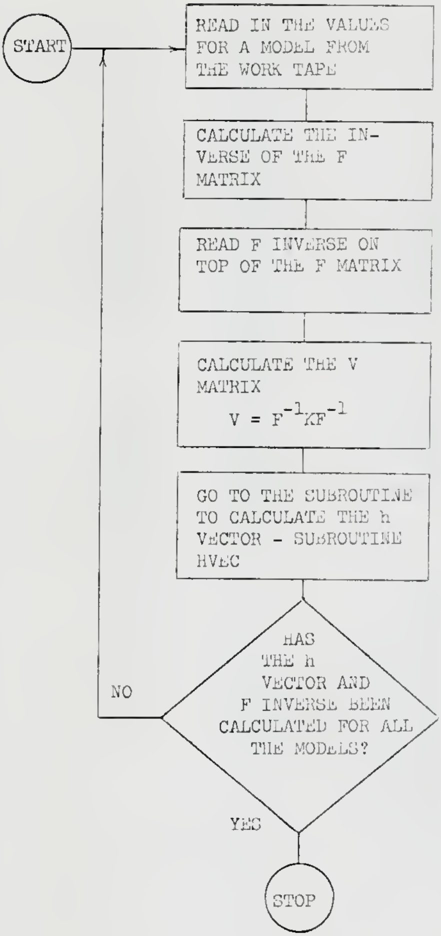

Fig. 3.4 Flow Diagram for "MATINV"

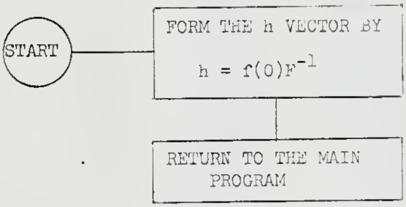

Fig. 3.5 Flow Diagram for Subroutine "HVEC"

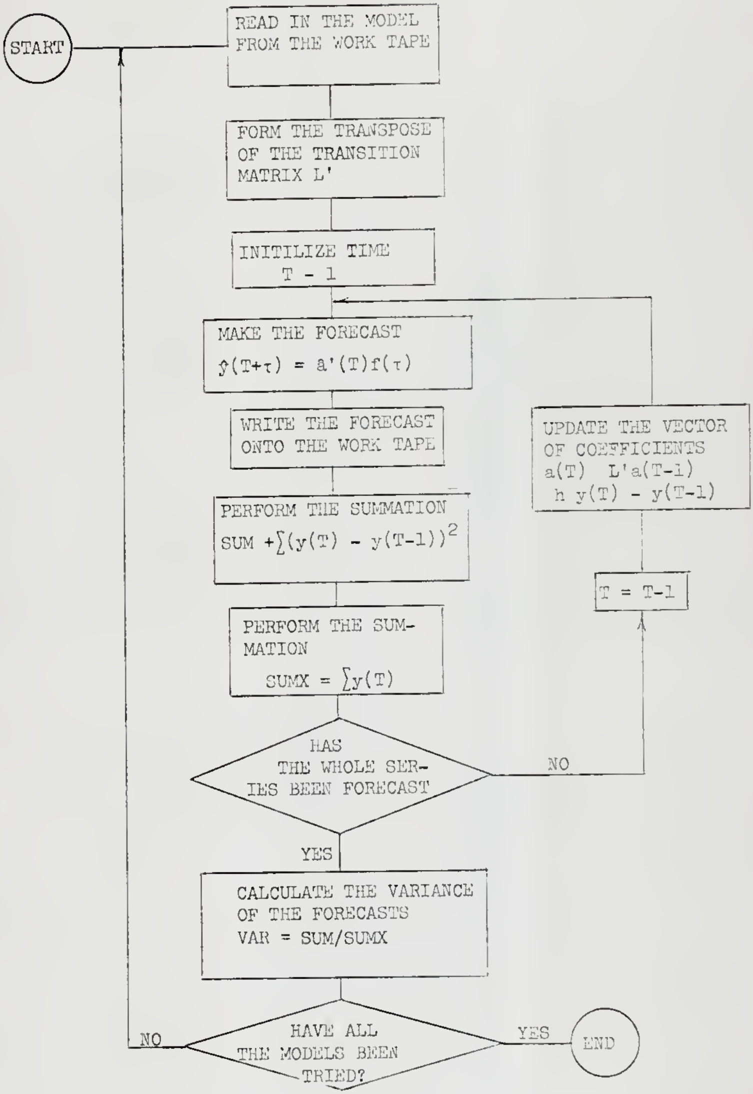

Fig. 3.6 Flow Diagram for "FCST"

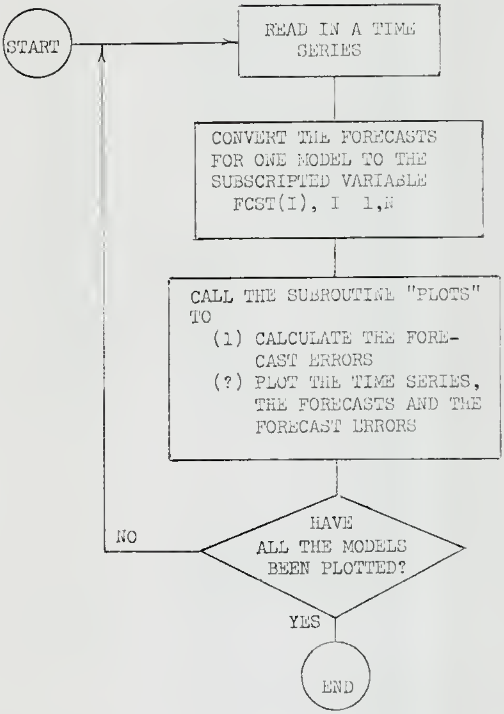

Fig. 3.7 Flow diagram for "PLOTTER"

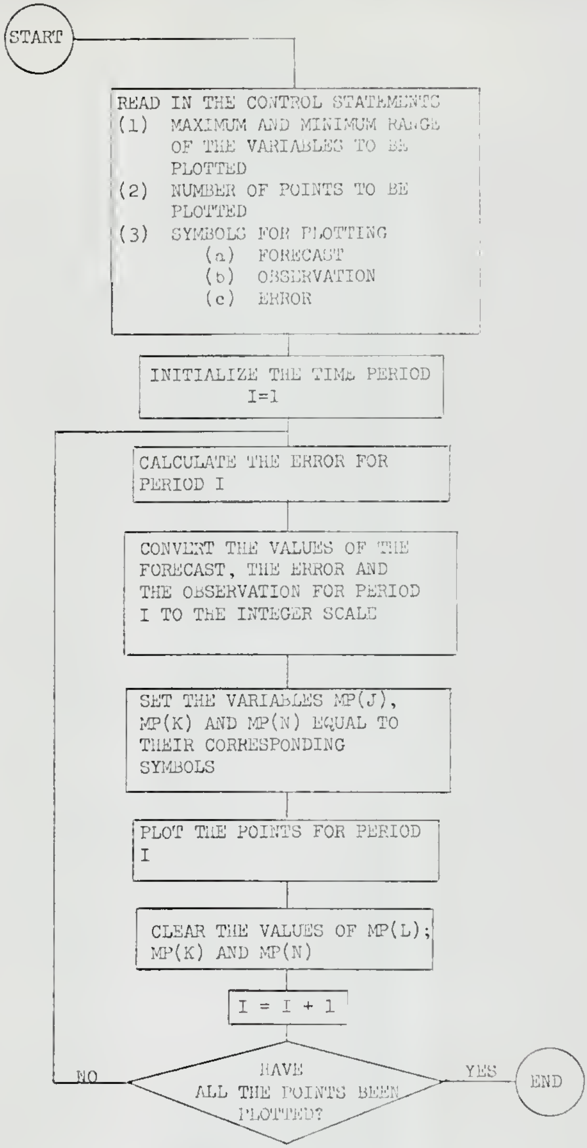

Fig. 3.8 Flow diagram for subroutine "PLOTS"

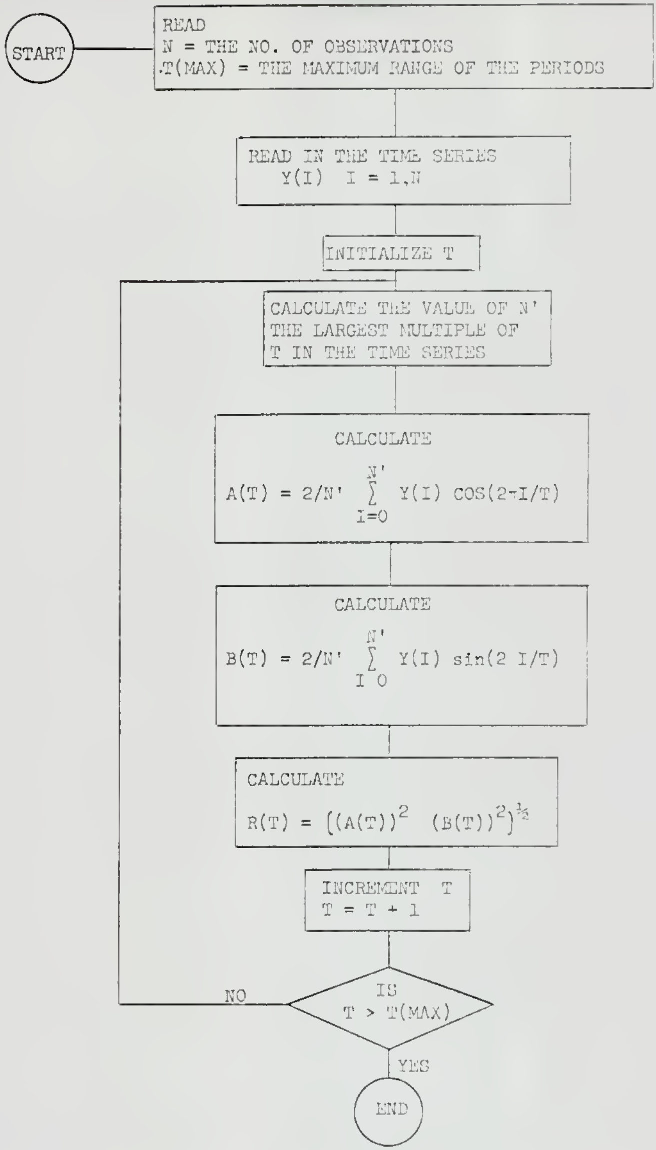

Fig. 3.9 Program to Calculate Power Spectrum

PHASE I INCONT

<u>WRITE ONTO WORK TAPE 7</u>

X(I)	= The Time Series
ND	= The Number of Observations
NTOU	= The Basic Period
N	= The Number of Fitting Functions
F(I)	= The Initial Value of the Vector of Fitting Functions
C(I)	= The Initial estimate of the Vector of Coefficients
MN	= The Model Number
CHK(I)	= The Change Vector
TM(I,J)	= The Transition Matrix

PHASE II RAY

<u>READ FROM WORK TAPE 7</u>

THE PHYSICAL RECORD WRITTEN IN PHASE I

<u>WRITE ONTO WORK TAPE 6</u>

N	= The Number of Fitting Functions
ND	= The Number of Observations
C(I)	= The Initial Values of the Vector of Coefficients
F1(I)	= The Value of the Vector of Fitting Functions Evaluated for One Period From the Time Origin
X(I)	= The Time Series
BETA	= The Value of the Smoothing Constant
MN	= The Model Number
TM(I,J)	= The Transition Matrix

<u>WRITE ONTO WORK TAPE 5</u>

N	= The Number of Fitting Functions
F(I)	= The Initial Value of the Vector of Fitting Functions
$(\mathrm{BETA})^{1/N}$	= The effective Value of the smoothing Constant
F(I,J)	= The F Matrix
K(I,J)	= The K Matrix

Fig. 3.10 Internal Data Transmission Between Phases

Fig. 3.10 CONTINUED

PHASE III MATINV

READ FROM WORK TAPE 5

THE PHYSICAL RECORD WRITTEN ON PHASE II

WRITE ONTO WORK TAPE 7

F = The Inverse of the F Matrix
h = The h Vector

PHASE IV FORECAST

READ FROM WORK TAPE 6

THE PHYSICAL RECORD WRITTEN IN PHASE III

READ FROM WORK TAPE 7

THE PHYSICAL RECORD WRITTEN IN PHASE III

WRITE ONTO WORK TAPE 5

ND = The Number of Observations
X(I) = The Time Series

PHASE V PLOTTER

READ FROM WORK TAPE 5

THE PHYSICAL RECORD WRITTEN IN PHASE IV

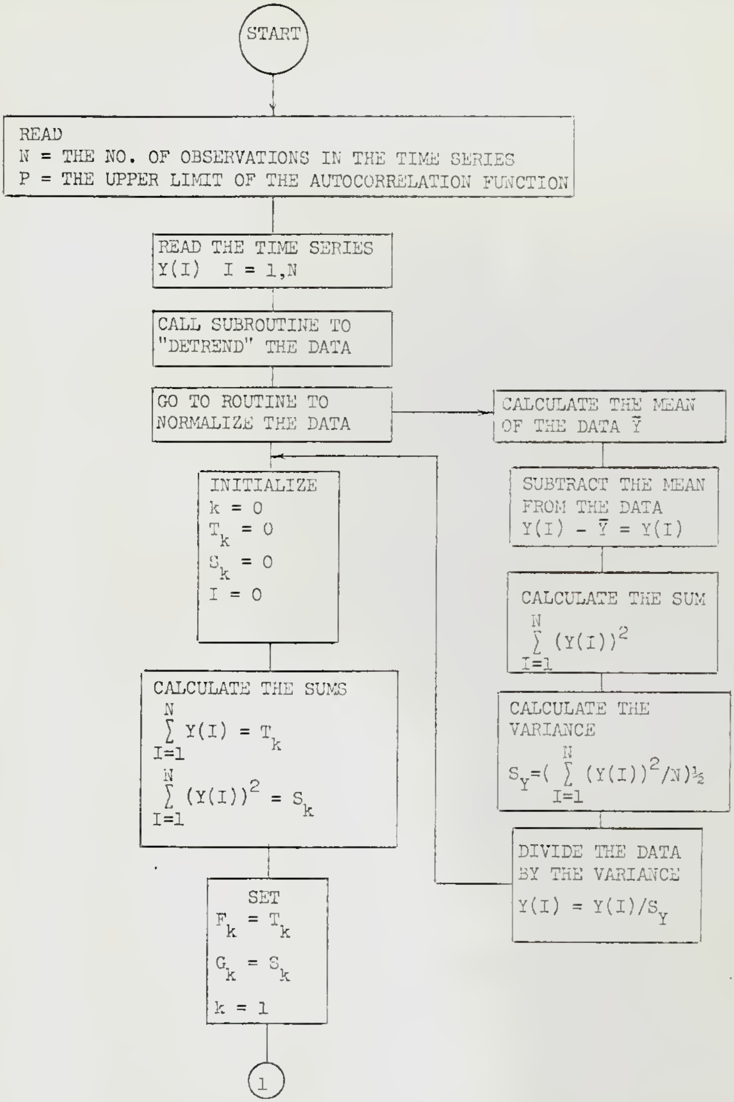

Fig. 3.11 Flow Diagram for Program to Calculate Autocorrelation Function

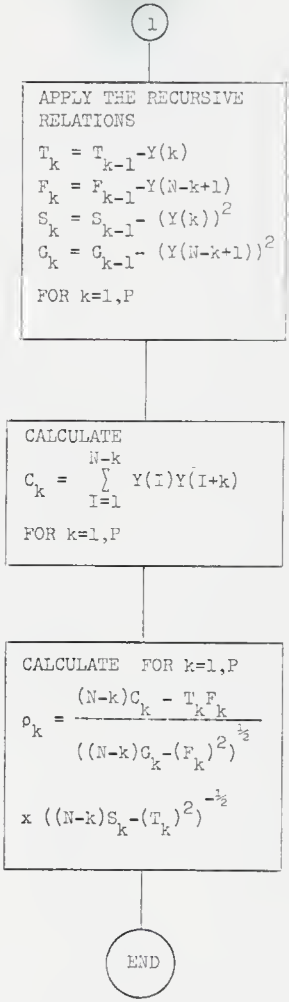

Fig. 3.11 continued

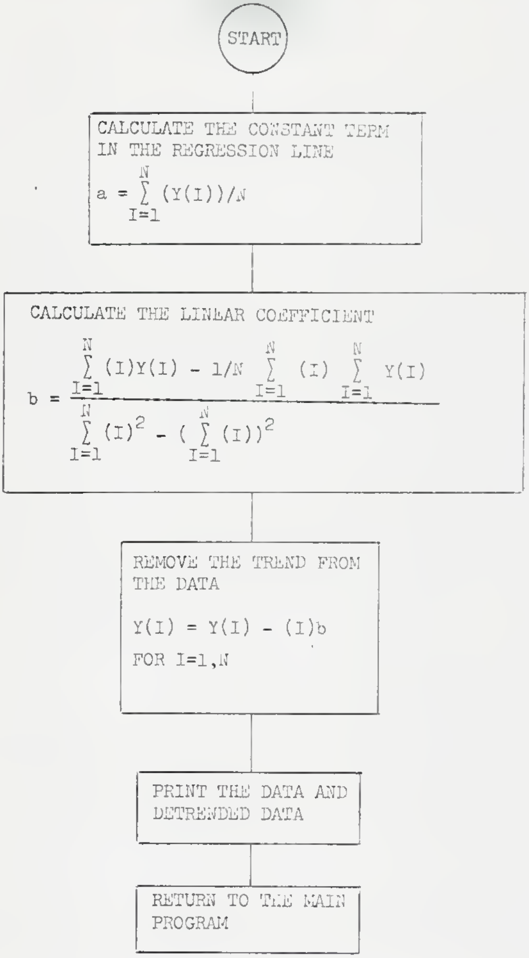

Fig. 3.12 Flow Diagram for Subroutine "TREND"

4 APPLICATIONS

The application of the analysis in the first three chapters is carried out on two time series. The first time series under consideration is the international airline data (Fig. 4.1) and the second time serics is the sheep production data (Fig. 4.2). The application takes the following form. First, "detrending" is performed to allow the data to be viewed without the effects of the secular trend of the mean. Second, autocorrelation analysis is performed (1) to determine if there is significant evidence that the time series is generated by a process and (2) to determine the basic period of the time series. Third, spectral analysis is applied to determine the contribution of the harmonics contained within the basic period. Fourth, the time series is represented by a model. Finally, general exponential smoothing is performed.

Within the context of the application of general exponential smoothing and the preceeding analysis the effectiveness and sensitivity of each technique is investigated. In the case of trend analysis the questions to be answered are:

(1) Does the "detrended" form of the data permit more effective analysis of the time series?

(2) Can trend analysis be used to effectively determine the trend of the mean of the time series?

(3) Can trend analysis be used to provide quantitative estimates for the trends of the periodic component of the time series?

As an aid to the trend analysis a Fortran program written for the I.B.M. 1410 computer is provided which plots the "detrended" data to make the results more useable.

In the case of the autocorrelation analysis the application to the two series provides answers to the questions

(1) Does the autocorrelation function give a true measure of the basic period of the process?

(2) Can the autocorrelation function be used as a test of significance that a process exists?

In the case of the international airline data (Fig. 4.1) it is quite evident that a process exists. Moreover, it is quite evident that the basic period of the periodic component of the time series is around twelve months. In this case autocorrelation analysis can be verified in view of the expected results. The sheep data (Fig. 4.2), however, is not quite as obvious. In this case the trend of the mean seems to be following a decreasing function, but there is no immediate indication of the basic period of the periodic component or, as a matter of fact, that a periodic component exists at all. In contrast to the international airline passenger data in which the results of the autocorrelation analysis can be immediately evaluated, the results of the analysis in the case of the sheep data can be evaluated only in the final stage of the analysis, the forecasting stage, when the effects of the choice of the basic period and the forecasting model can be tested.

Using spectral analysis it is determined in section 1.4 that the contribution of each harmonic within the basic period can be measured. This analysis, it is proposed, will lead to the most optimal selection of the periodic terms to include in the forecasting model. However, no evidence is given in section 1.4 that the application of spectral analysis to an actual time series will lead to the obvious distinction between the contribution of the harmonics. That is, it is not yet shown that the application of spectral analysis yields results which justify the selection of a limited number of harmonics to

adequately describe the periodic component of the time series.

In section 1.5 the effect of a linear trend in the mean of the time series on spectral analysis is considered. This analysis provides sufficient evidence that the "detrended" form of the data should be used for the spectral analysis. The application of spectral analysis to the "detrended form of the two time series under consideration investigates the effectiveness of this method.

The final point to be investigated in spectral analysis is its relation to autocorrelation analysis. In section 1.3 it is shown that the autocorrelation function has a local maximum at ω_i, the periods of the harmonics which describe the periodic component of the time series. Hence, if spectral analysis is carried out over a wide enough range the basic period of the time series as indicated by autocorrelation analysis should agree with the first harmonic of the series as indicated by spectral analysis

In summary, then, the application of spectral analysis investigates the following questions:

(1) Can spectral analysis be effectively used to determine the periodic terms to be included in the forecasting model?

(2) Is the use of the "detrended" form of the data an effective method of spectral analysis?

(3) Do the results of spectral analysis verify the results of autocorrelation analysis?

Continuing to the final stage, the analysis of general exponential smoothing which forms the contents of Chapter 2 presupposes the following:

(1) An adequate choice of the fitting function which make up the forecasting model can be determined.

(2) The basic period of the forecasting model can be adequately determined.

(3) A proper choice of the smoothing constant can be determined.

In the application of general exponential smoothing, trend analysis, autocorrelation analysis and spectral analysis can be effectively used to help determine (1) and (2). It is well recognized, however, that the results obtained by these measures are only approximations. Therefore, the effects of errors in these parameters should be investigated. This investigation takes the form of sensitivity analysis.

Furthermore, no method is available in the existing literature for determining the optimal value to the smoothing constant. As a matter of fact, no evidence is available to substantiate the existence of a singular optimal value of this constant. This research does not attempt to find this optimal. It is recognized, however, that a local value of the smoothing constant might exist for each particular time series and set of fitting functions. Hence, a parametric investigation of the smoothing constant is carried out for each time series.

The application of general exponential smoothing then, investigates the following questions:

(1) How sensitive is general exponential smoothing to the choice of fitting functions used in the forecasting model?

(2) How sensitive is general exponential smoothing to the choice of the basic period in the forecasting model?

(3) How sensitive is general exponential smoothing to the choice of the smoothing constant?

4.1 The Application of Trend Analysis

In the case of the international airline passenger data (Fig. 4.1) a linear trend seems to best represent the mean of this time series. The flow diagram for the computer program to "detrend" the time series is shown in Fig. 3.12 and the program itself is shown in Appendix E. The results of this program give a value of 112.0 for the constant term in the regression model and a value of 2.6 for the coefficient of the linear term. The detrended data is given in Fig. 4.3. Line A-A which is given as y = 112.0 suggests that the "detrending" has successfully removed the secular trend of the mean from the data.

The vertical dashed lines in Fig. 4.3 are constructed at twelve month intervals in order to indicate the basic period of the data. The most important information gleaned from the trend analysis of the international airline data, however, is the trend of the periodic component. Line B-B connects the peaks of the periodic component. Since the secular trend of the mean has been removed from the data the slope of the line B-B indicates the rate of growth of the amplitude of the periodic component.

Figure 4.4 is the plot of the "detrended" sheep data. In this case the linear trend is again assumed and the results yield a constant term of 2207.0 and a value of -12.2 for the coefficient of the linear term in the regression model. The line A'-A' which can be represented as y = 2207.0 also indicates that the linear model is a good representation of the trend of the mean. The "detrended" form of the sheep data does not obviously yield any further information at this point.

4.2 Application of Autocorrelation Analysis

The flow diagram for the computer program for autocorrelation analysis is given in Fig. 3.11 and the computer program is given in Appendix E. The autocorrelation function for the international airline passenger data is given in Table 4.1 and the plot of this function is given in Fig. 4.5. It can be seen that this function reaches its maximum at 12 periods and multiples thereof. Thus, the autocorrelation analysis bares out the results expected from the trend analysis. The peak of the autocorrelation function is a maximum at 12 months and decreases at multiples of this period. This decrease, or decay, in the maximum values of the autocorrelation function is due to the trend in the amplitude of the periodic component and further serves to verify the effectiveness of autocorrelation analysis.

In the case of the sheep data the autocorrelation function is given in Table 4.2 and the plot of this function is given in Fig 4.6. This function reaches a local maximum at 25 and again at 40 months. The maximum value at 40 months, however, is much more predominent. The range of the autocorrelation function for the sheep data is taken as 50 months. This range is not extended because the available range of the data is only 73 months.

4.3 Application of Spectral Analysis

The application of spectral analysis is performed on the "detrended" data. Hence, any distortion of the analysis due to the secular trend of the mean is eliminated. The computer program for the application of this analysis is given in Appendix E and the flow diagram is given in Fig. 3.13. As proposed earlier the purpose of spectral analysis is to determine the contributing frequencies in the time series. The plot of the power spectrum for the international airline data is given in Fig. 4.7 and the power spectrum is

given in Table 4.3. The power spectrum for this series has a clear maximum at both 6 and 12 months. This analysis indicates a 12 month period with a harmonic. Moreover, since one of the purposes of spectral analysis is to verify the results of autocorrelation analysis this basic period of 12 months obtained is quite significant. The clear distinction of the 6 month harmonic indicated by spectral analysis suggests a more thorough analysis of the periocity of the time series using this method.

In the case of the sheep data another interesting result is obtained. Figure 4.8 and Table 4.4 give the power spectrum and the plot of the power spectrum for this time series. The power spectrum reaches a local maximum at 18 months and a maximum at 36 months. In this case, however, the results do not agree with those obtained by autocorrelation analysis which indicates a basic period of 40 months. At this point no conclusion can be drawn about the validity of the two techniques. The discussion of the comparison of the two methods must be curtailed until the effectiveness of the forecasting models can be considered.

4.4 Application of Exponential Smoothing

The flow diagram for the computer program used in exponential smoothing is given in Fig. 3.1 through 3.7 and the program is given in Appendix E. Using this program the results of trend analysis, autocorrelation analysis and spectral analysis can be investigated. Moreover, the effect of the value of the smoothing constant on the forecast can be investigated.

The application of general exponential smoothing is aimed at determining the parameters which influence the forecast and the sensitivity of the forecast to those parameters. The parameters chosen for investigation are

(1) The fitting functions used to describe the process

(2) The basic period of the forecasting model

(3) The smoothing constant

In the comparison of forecasts a measure of the effectiveness of each forecast must be provided. Brown (1,393) suggests the following measure of the effectiveness of the forecast

$$\sum_{t=1}^{T}(y_t-\hat{y}_t)^2/(T-n)$$

where y_t is the observation at time t

$\hat{y}_t$ is the forecast for time t

T is the total number of observations in the time series

n is the number of fitting functions in the forecasting model

Upon consideration of this measure of effectivenss proposed by Brown the following objection is incurred. Although the above measure of effectiveness is useful for comparing forecasts of the same time series it cannot adequately compare forecasts between different time series. The reason for this inadequacy lies in the fact that as the size of the observations in the time series increases, the size of the term above increases even though the errors may not be proportionately as large. Hence, a new measure of the effectiveness of the forecast is devised as

$$\sum_{t=1}^{T}(y_t-\hat{y}_t)^2/\sum_{t=1}^{T}y_t \tag{4.4.1}$$

with this new measure of effectiveness the forecasts for different time series can be compared on an equal basis.

The application of general exponential smoothing begins with the construction of the forecasting model (see section 1.1). This model represents the process which generates the time series. In the selection of the fitting functions used to represent the process the results of trend analysis, autocorrelation analysis and spectral analysis are utilized. Beginning with the international airline data two models are chosen. The first model represents a growing sinusoid with a basic period of 12 months and the second model represents a growing sinusoid with a basic period of 12 months plus a harmonic at 6 months.

The mathematical representation of the first model is

$$y(T+\tau) = (a_1+a_2t) + (a_3+a_5t)\sin(2\Pi t/12) + (a_4+a_6t)\cos(2\Pi t/12)$$

where the terms:

$(a_1 + a_2t)$ represent the linear trend of the mean

$a_3 \sin(2\Pi t/12) + a_4 \cos(2\Pi t/12)$ represent the 12 month periodic component

$(a_5t)\sin(2\Pi t/12) + (a_6t)\cos(2\Pi t/12)$ represent the growing amplitude of the periodic component.

The last set of fitting functions also give the model the ability to adapt to shifting phase angles. In the case of the second model the only difference in the fitting functions is that the terms

$$a_7 \sin(4\Pi t/12) + a_8 \cos(4\Pi t/12)$$

are included to represent the harmonic at 6 months.

The initial estimates of the coefficients for the terms which represent the linear trend of the mean are obtained from the regression analysis as $a_1 = 112.0$ and $a_2 = 2.6$. In the case of the coefficients a_5 and a_6 of the terms which represent the growth in the amplitude of the periodic component trend analysis is used. In section 4.1 these coefficients are estimated through the use of Fig. 4.3 as 1.9. The initial value of the coefficients of the periodic terms a_3, a_4, a_7 and a_8 are estimated by the method proposed by Brown (1,194).

$$a_k = 2/P \sum_{k=1}^{P} y_k \sin \omega(k)$$

for the coefficients a_3 and a_7 and

$$a_k = 2/P \sum_{k=1}^{P} y_k \cos \omega(k)$$

for the coefficients a_4 and a_8.

The purpose of this investigation is to determine the increase in effectiveness of the forecast due to the inclusion of the harmonic term as indicated by the spectral analysis, and the effect of the value of the smoothing constant on the forecasts. The results of this application are shown in Table 4.5. The results are reported according to the forecasting model, the value of the smoothing constant, β and the basic period, P, of the forecasting model. The measure of effectiveness given for each forecast is the variance of the forecast (4.4.1).

In this application the effective value of the smoothing constant is reported. That is,

$$\beta^n = \beta_{(\text{effective})}$$

where n is the number of degrees of freedom in the model. Along with the results in Fig. 4.5 the other results obtained for each forecast are given

in Table 4.6. This table gives the following values for each forecast

EFFECTIVE BETA

F INVERSE

h VECTOR

VARIANCE OF THE COEFFICIENTS

The advantage of tabulating these results is that once the h vector is calculated for one set of fitting functions and a particular value of the smoothing constant then, since the value of this vector is independent of the time series, it can be used to forecast any time series where the same fitting functions and smoothing constant are used.

The results of Table 4.5 clearly indicate that the value of the smoothing constant and the choice of the fitting functions are significant parameters in the forecast. In the case of this data a value for the smoothing constant of 0.70 is far superior to a value of 0.90 for either of the two models. However, it is interesting to note that the growing sinusoidal model fares better for a value of the discount factor of 0.90 than does the harmonic model. When the discount factor is reduced to 0.70 the situation is reversed. The most important information obtained from these results is that the best forecast is obtained using the harmonic model. This varifies the results of spectral analysis.

The effect on the trace of the forecast for a change in the smoothing constant can be seen in Figs. 4.9 and 4.10. These figures clearly indicate that the peaks of the periodic component of the forecast are cut down as the value of the smoothing constant is increased. This analysis of the smoothing of the peaks of the forecast is one of the most important outcomes of this investigation.

In Fig. 4.9 the ability of the model to adjust to a phase change can be clearly noted. At the beginning of the forecast the trace is clearly out of phase with the observations. However, at the end of the time series the trace is exactly in phase.

In the case of the sheep data three forecasting models are used

(1) Linear

(2) Linear plus sinusoid

(3) Linear plus sinusoid plus harmonic

Along with the investigation of the choice of fitting functions; and the value of the smoothing constant, in this model the choice of the basic period is also investigated. The investigation of the basic period is noteworthy in this case because of the results of the autocorrelation analysis and spectral analysis noted in section 4.3.

The mathematical representation of the models used in forecasting this time series are

LINEAR

$$\xi(T + t) = a_1 + a_2 t$$

LINEAR WITH SUPERIMPOSED SINUSOID

$$\xi(T + t) = a_1 + a_2 t + a_3 \sin(2\Pi t/P) + a_4 \cos(2\Pi t/P)$$

LINEAR MODEL WITH SUPERIMPOSED SINUSOID AND HARMONIC

$$\xi(T + \tau) = a_1 + a_2 t + a_3 \sin(2\Pi t/P) + a_4 \cos(2\Pi t/P) + a_5 \sin(4\Pi t/P) + a_6 \cos(4\Pi t/P)$$

where P is the basic period of the forecasting model and τ is the forecast period. In contrast to the models used in the international airline passenger data these models are presented in a more general manner to allow for the parametric study of the choice of the basic period. Table 4.7 gives the results of this study.

Just as in the case of the forecasts for the international airline passenger data these results indicate that the value of the discount factor and the choice of the fitting functions are significant parameters in the forecast. In this case the effectiveness of all the models used increases as the discount factor decreases.

The most interesting study in this time series, however, is the study of the choice of the basic period. The significance of choosing the correct basic period can be seen from Table 4.7 which indicates that if the basic period is chosen incorrectly at 12 months then the linear model is more effective. Hence, an incorrect choice of the basic period negates the effect of the periodic terms in describing the time series and even makes the inclusion of these terms, at the cost of increasing the calculations, detrimental to the forecast.

With the results of the application of exponential smoothing available the results of autocorrelation analysis and spectral analysis on the sheep data can be reinvestigated. Although the basic period of 36 given by spectral analysis is more effective than the result of 40 given by autocorrelation analysis the harmonic model suggested by spectral analysis is less effective

than the basic sinusoidal model. These discrepancies, however, can be partially explained by the range of the available data. The international airline passenger data in which the results of the analysis are compatible contains 12 periods of data while the sheep data contains barely 2. Hence, it is felt that the results of this study indicate one more criteria for the use of general exponential smoothing. The data should cover enough basic periods to allow the smoothing technique to become effective.

PERIOD	R
1	.73
2	.29
3	-.08
4	-.33
5	-.41
6	-.45
7	-.44
8	-.37
9	-.12
10	.23
11	.66
12	.92
13	.67
14	.23
15	-.14
16	-.37
17	-.45
18	-.5[illegible]
19	-.51
20	-.45
21	-.19
22	.18
23	.61
24	.88
25	.64
26	.21
27	-.15
28	-.37
29	-.46
30	-.5[illegible]
31	-.5[illegible]
32	-.46
33	-.21
34	.16
35	.59
36	.86
37	.65
38	.23
39	-.0[illegible]
40	-.3[illegible]

FIG. 4.1 AUTOCORRELATION FUNCTION FOR AIRLINE DATA

41	-.41
42	-.48
43	-.53
44	-.5[illegible]
45	-.25
46	.13
47	.55
48	.82
49	.62
50	.22
51	-.08
52	-.28
53	-.39
54	-.49
55	-.55
56	-.56
57	-.32
58	.05
59	.44
60	.71
61	.5[illegible]
62	.13
63	-.14
64	-.32
65	-.42
66	-.52
67	-.59
68	-.6[illegible]
69	-.36
70	.0[illegible]
71	.4[illegible]
72	.67
73	.49
74	.14
75	-.11
76	-.26
77	-.35
78	-.46
79	-.55
80	-.56

FIG. 4.1 CONTINUED

PERIOD	R
1	.82
2	.53
3	.29
4	.15
5	.11
6	.09
7	.02
8	-.09
9	-.23
10	-.37
11	-.49
12	-.54
13	-.51
14	-.43
15	-.32
16	-.21
17	-.1[illegible]
18	-.11
19	-.18
20	-.22
21	-.16
22	-.0
23	.19
24	.1
25	.02
26	-.08
27	-.17
28	-.23
29	-.22
30	-.12
31	.01
32	.14
33	.26
34	.33
35	.35
36	.35
37	.36
38	.43
39	.58
40	.72
41	.67
42	.48
43	.19
44	-.12

~~FIG.~~ Table 4.2 AUTOCORRELATION FUNCTION FOR SHEEP DATA

A(I)	B(I)	R(I)	I
-2.6909600	-.1345205	2.6943200	5
19.3717580	16.136046	25.2118380	6
.7956058	-.7461768	1.0907650	7
-.5610496	5.4535620	5.4823447	8
.1521238	3.5783589	3.5815901	9
-4.5237609	-4.9677057	6.7188168	10
-9.5527615	9.6671011	13.5871770	11
-45.4427770	5.3682407	45.7587510	12
7.9891889	-15.496477	17.4346720	13
9.1534579	4.7859903	10.3291550	14
-5.9185971	.6573547	5.9549891	15
-.7482543	-1.6923416	1.8503795	16
-1.4585645	3.364016	3.6666072	17
.6890328	-4.1212851	4.1784864	18
-2.6352047	-1.3322884	2.9528448	19
3.7815674	3.2022389	4.9552572	20
.4096793	4.5957660	4.6139890	21
.4102819	-.4917886	.6404587	22
-1.5726409	4.5032335	4.7699374	23
1.4384364	2.1948391	2.6241983	24
2.9611672	5.8909796	6.5933401	25
8.8604870	-.085210	8.8608949	26
1.6408514	-2.2574280	2.7907653	27
4.5932860	-2.6047976	5.2804576	28
3.5356237	1.3538123	3.7859523	29
-.1470238	1.1067774	1.1164999	30
-1.9255327	3.4905975	3.9864698	31
4.5081015	3.2721962	5.5704786	32
2.6152783	.8205563	2.7409834	33
.0018217	.8349539	.8349559	34
4.6276560	-.2176798	4.6327719	35
.7363869	-1.7951662	1.9403316	36
-1.3050109	5.0298174	5.1963553	37
-.6666127	6.8219076	6.8543985	38
4.2487874	6.494764	7.7610650	39
1.9263122	6.5405445	6.8183123	40

TABLE 4.3   POWER SPECTRUM FOR AIRLINE DATA

A(I)	B(I)	R(I)	I
1.2407496	7.5337514	7.6352377	41
3.5765891	8.1096308	8.8632992	42
9.6752679	6.2404171	11.5131900	43
8.3841776	4.2423931	9.3963984	44
7.1160619	2.9418872	7.7001959	45
8.7785713	1.5602106	8.9161396	46
13.1607560	-1.6039073	13.2581290	47
9.7507890	-4.2358422	10.6310960	48
5.8010021	-1.7388835	6.0560161	49
4.9174372	-1.9962514	5.3071834	50
5.2918680	-2.2479780	5.7495441	51
8.3214625	-3.2275624	8.9254619	52
7.3868407	-4.3244308	8.5595615	53
4.1026434	-4.7237444	6.2566310	54
1.0449346	-4.5242634	4.6433648	55
-1.0617951	-3.8699448	4.0129634	56
-1.5733583	-3.0260649	3.4106526	57
1.2644329	-2.7724021	3.0471300	58
-.1199097	-2.5650119	2.5678130	59
-3.3816608	-1.7659442	3.8149943	60
-5.7564987	-.5327325	5.7810959	61
-6.3925674	.8460000	6.4483035	62
-5.5354076	2.1018292	5.9210143	63
-1.3925134	2.6077264	2.9562352	64
-1.1883522	2.8096931	3.0506643	65
-2.7014901	3.2863276	4.2541734	66
-3.6644789	3.9108856	5.3594233	67
-3.6179184	4.6102800	5.8603754	68
-1.8224397	5.0665305	5.3843299	69
2.9014497	4.7558034	5.5710022	70
3.5709085	4.1628157	5.4845611	71
1.9256423	3.8577435	4.3116440	72
-.3685073	11.7221230	11.7279110	73
-1.1416073	12.2012690	12.2545570	74
-1.0222077	12.6547540	12.6959700	75
-.8902031	13.0831700	13.1134180	76
-.7743109	13.4893650	13.5115690	77
.4487438	13.7854060	13.7927040	78
2.8258414	13.884882	14.1695180	79

TABLE 4.3 (CONTINUED)

A(I)	B(I)	R(I)	I
9.9500209	6.6609684	11.9737770	5
35.0316210	7.3195859	35.7881330	6
-13.3377480	32.3318270	34.9748750	7
32.4712490	66.307059	73.8309310	8
-3.3339637	-24.290088	24.5178190	9
33.0320220	-.8567371	33.0431230	10
6.8349347	-25.212314	26.1223600	11
13.4308350	-20.745791	24.7138630	12
-23.1792910	-43.043977	48.8882640	13
-38.1888220	-2.1998473	38.2521230	14
7.0530832	10.7685839	12.8727740	15
-12.9974900	-31.479678	34.0573730	16
-56.8663860	-29.208660	63.9291070	17
-80.7530060	27.661924	85.3593990	18
-42.5203850	121.222650	128.4636400	19
21.0487260	125.6580000	127.4087000	20
70.1836930	86.4004580	111.3139000	21
91.9226860	31.607639	97.2050410	22
74.4070800	-17.927160	76.5362400	23
39.6314820	-40.902037	56.9528740	24
39.7346920	36.352436	53.8548440	25
32.5516880	35.891007	48.4538530	26
11.6821440	44.3637250	45.8760440	27
-3.6673356	61.363527	61.4730060	28
-3.2556033	80.1263290	80.1924250	29
18.8103930	91.3476690	93.2642680	30
45.1305050	91.8786660	102.3642900	31
66.7364460	84.638762	107.7843600	32
96.1385960	68.350432	117.9593500	33
113.9964500	45.383657	122.6982400	34
120.8952900	21.192619	122.7387200	35
125.0038700	-2.8807288	125.0370400	36
90.6577020	-16.123709	92.0803400	37
84.0657820	-20.0148920	86.4155700	38
78.0725890	-23.0554910	81.4056650	39
73.7369100	-25.585575	78.0496730	40

TABLE 4.4 POWER SPECTRUM FOR SHEEP DATA

Table 4.5 The Results of the Forecasts of the International Airline Passenger Data Using General Exponential Smoothing

MODEL	GROWING SINUSOID P = 12
BETA	VARIANCE OF FORECASTS
0.70	2.977
0.90	8.774
MODEL	GROWING SINUSOID WITH HARMONIC P = 12
0.70	1.681
0.90	13.343

Linear

$1-\alpha = 0.75$

```
                    BETA=   .866025

F INVERSE
          .250000              .017949
          .017949              .002776

THE H VECTOR IS
          .250000
          .017949

     THE VARIANCE OF THE COEFFICIENTS IS
          .169367
          .000740
```

Table 4.6

Linear

$1-\alpha = 0.90$

BETA= .949683

F INVERSE

.100001	.002633
.002633	.000142

THE H VECTOR IS
.100001
.002633

THE VARIANCE OF THE COEFFICIENTS IS
.064458
.000036

Table 4.6 continued

Linear

1-α = 0.95

```
                          BETA=   .974679

F  INVERSE
               .050001                 .000641
               .000641                 .00001

THE  H  VECTOR  IS
               .050001
               .000641

            THE  VARIANCE  OF  THE  COEFFICIENTS  IS
               .031729
               .000004
```

Table 4.6 continued

Linear & Sinusoid

$1-\alpha = 0.75$ $P = 12$

BETA= .930604

F INVERSE

.144588	.005200	.036607	[illegible]
.005200	.000373	.001285	[illegible]
.036607	.001295	.154317	[illegible]
-.015090	-.000632	.00454[illegible]	[illegible]

THE H VECTOR IS
.129498
.004568
.041148
.120502

THE VARIANCE OF THE COEFFICIENTS IS
.035101
.000086
.068774
.070726

Table 4.6 continued

Linear & Sinusoid

$1-\alpha = 0.90$ $P = 12$

BETA= .974003

F INVERSE

.051851	.000682	.00504[illegible]	-.001[illegible]
.000682	.000019	.0000[illegible]	-.000[illegible]
.005040	.000066	.05323[illegible]	.001014
-.001608	-.000022	.001014	.0513[illegible]7

THE H VECTOR IS
.050242
.000659
.006054
.049758

THE VARIANCE OF THE COEFFICIENTS IS
.032447
.000004
.026124
.026316

Table 4.6 CONTINUED

Linear & Sinusoid

$1-\alpha = 0.75$ $P = 36$

BETA= .3306[illegible]4

F INVERSE

.245097	.008591	.13[illegible]152	-.[illegible]7790[illegible]
.008591	.000431	.004[illegible]	-.003345
.138152	.0041[illegible]	.[illegible]3[illegible]36	-.143[illegible]
-.077908	-.003345	-.143[illegible]	.1[illegible]719

THE H VECTOR IS

.167188
.005245
.123792
.082811

THE VARIANCE OF THE COEFFICIENTS IS

.088191
.000073
.075650
.056792

Table 9.6 continued

Linear & Sinusoid

$1-\alpha = 0.95 \quad P = 36$

BETA= .98725[illegible]

F INVERSE

.025873	.000165	.003736	-.000604
.000165	.000052	.000023	-.000064
.003736	.000023	.028254	.00[illegible]34
-.000604	-.000064	.007834	.[illegible]5334

THE H VECTOR IS
.025269
.000161
.004570
.024731

THE VARIANCE OF THE COEFFICIENTS IS
.015817
.000000
.012589
.012823

Table 4.6 continued

Growing Sinusoid

$1-\alpha = 0.70$ $P = 12$

BETA= .942285

C INVERSE

.124	.003	.051	-.019	.001	-.000
.003	.000	.001	-.000	.000	-.000
.051	.001	.264	.013	.007	.000
-.019	-.000	.013	.214	.000	.006
.001	.000	.007	.000	.000	.000
-.000	-.000	.001	.004	.000	.000

THE H VECTOR IS
.104834
.003036
.065137
.195165
.001913
.005701

THE VARIANCE OF THE COEFFICIENTS IS
.071303
.000048
.146988
.146346
.000098
.000102

Table 4.6 continued

Growing Sinusoid

$1-\alpha = 0.90$ $P = 12$

BETA=   .982593

INVERSE

.034	.000	.004	-.001	.000	-.000
.000	.000	.000	-.000	.000	-.000
.004	.000	.070	.001	.000	.000
-.001	-.000	.001	.067	.000	.000
.000	.000	.000	.000	.000	.000
-.000	-.000	.000	.000	.000	.000

THE H VECTOR IS
.033473
.000293
.006398
.066526
.000056
.000583

THE VARIANCE OF THE COEFFICIENTS IS
.021756
.000001
.043806
.043278
.000002
.000002

Table 4.6 CONTINUED...

Growing Sinusoid & harmonic

$1-\alpha = 0.70$ $P = 12$

BETA= .956394

INVERSE

.091	.002	.029	-.008	.000	-.000	.005	-.005
.002	.000	.000	-.000	.000	-.000	.000	-.000
.029	.000	.191	.011	.004	.000	.001	-.012
-.008	-.000	.011	.168	.000	.003	.016	-.009
.000	.000	.004	.000	.000	.000	.000	-.000
-.000	-.000	.000	.003	.000	.000	.000	-.000
.005	.000	.001	.016	.000	.000	.091	-.000
-.005	-.000	-.012	-.009	-.000	-.000	-.000	.096

THE H VECTOR IS
.077159
.001691
.027763
.150309
.000611
.003316
.022328
.072031

THE VARIANCE OF THE COEFFICIENTS IS
.053980
.000021
.109720
.106372
.000041
.000042
.043929
.043850

Table 4.6 continued

Growing Sinusoid & harmonic

$1-\alpha = 0.90 \quad P = 12$

BETA= .986916

INVERSE

.026	.000	.002	-.000	.000	-.000	.000	-.000
.000	.000	.000	-.000	.000	-.000	.000	-.000
.002	.000	.053	.001	.000	.000	.000	-.000
-.000	-.000	.001	.051	.000	.000	.001	-.000
.000	.000	.000	.000	.000	.000	.000	-.000
-.000	-.000	.000	.000	.000	.000	.000	-.000
.000	.000	.000	.001	.000	.000	.025	.000
-.000	-.000	-.000	-.000	-.000	-.000	.000	.126

THE H VECTOR IS
.025062
.00 164
.002701
.050024
.000017
.000329
.002221
.024913

THE VARIANCE OF THE COEFFICIENTS IS
.016354
.000000
.032872
.032549
.000001
.000001
.013146
.013155

Table 4.6 CONTINUED

Table 4.7 The Results of the Forecasts of the Sheep Data Using General Exponential Smoothing

LINEAR MODEL

BETA	VARIANCE OF FORECASTS
0.75	9.025
0.90	11.417

LINEAR MODEL WITH SUPERIMPOSED SINUSOID

P = 12

0.75	11.004
0.90	13.212

P = 36

0.70	7.550
0.75	7.644
0.90	9.352
0.95	12.295

P = 40

0.70	9.393

LINEAR MODEL WITH SUPERIMPOSED SINUSOID AND HARMONIC

P = 36

0.70	10.196

P = 40

0.70	[illegible]

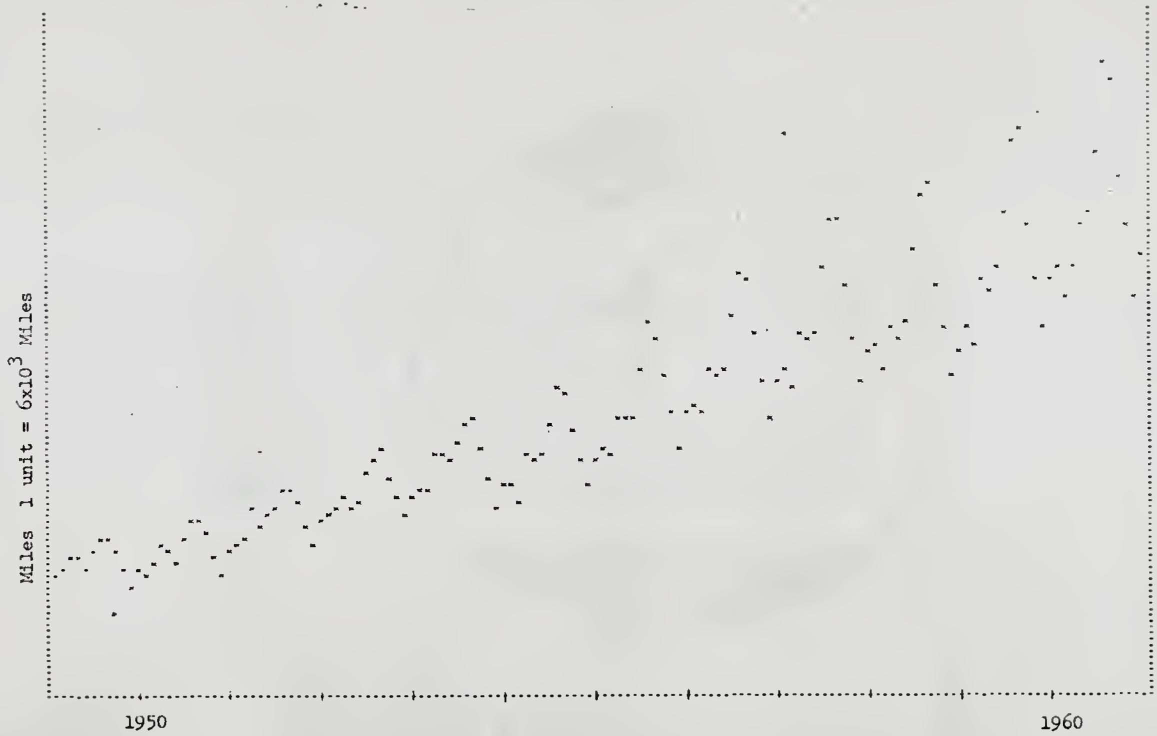

Fig. 4.1 International Airline Data

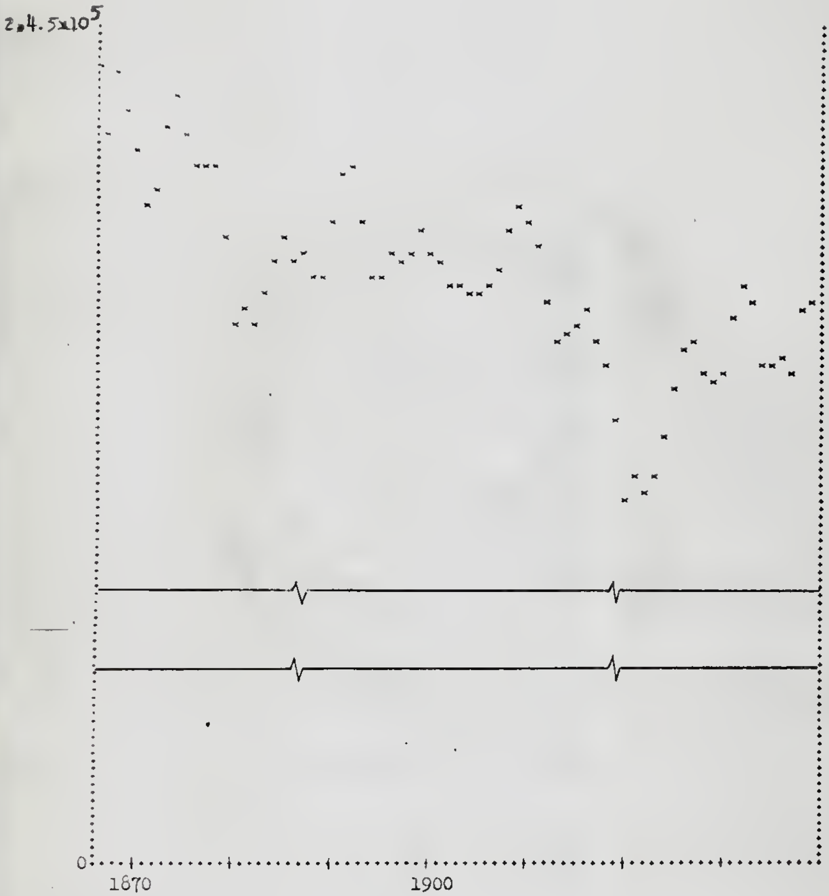

Fig. 42 Sheep Population in England and Wales

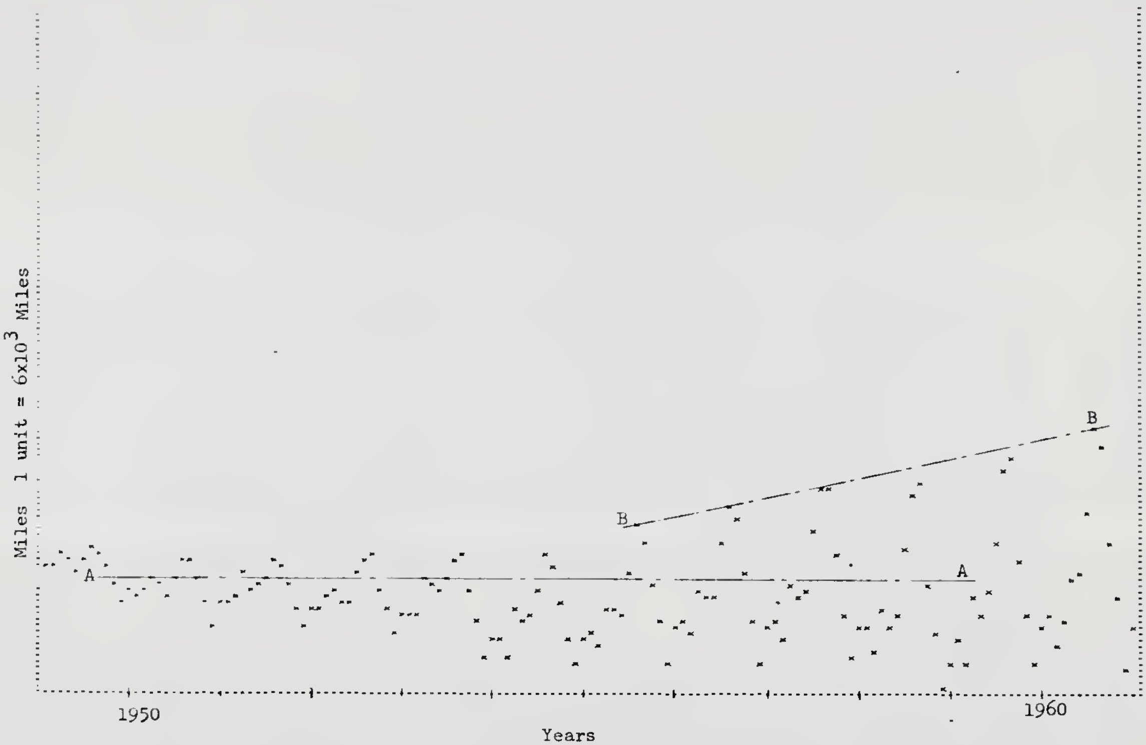

Fig. 4.3 "Detrended" International Airline Data

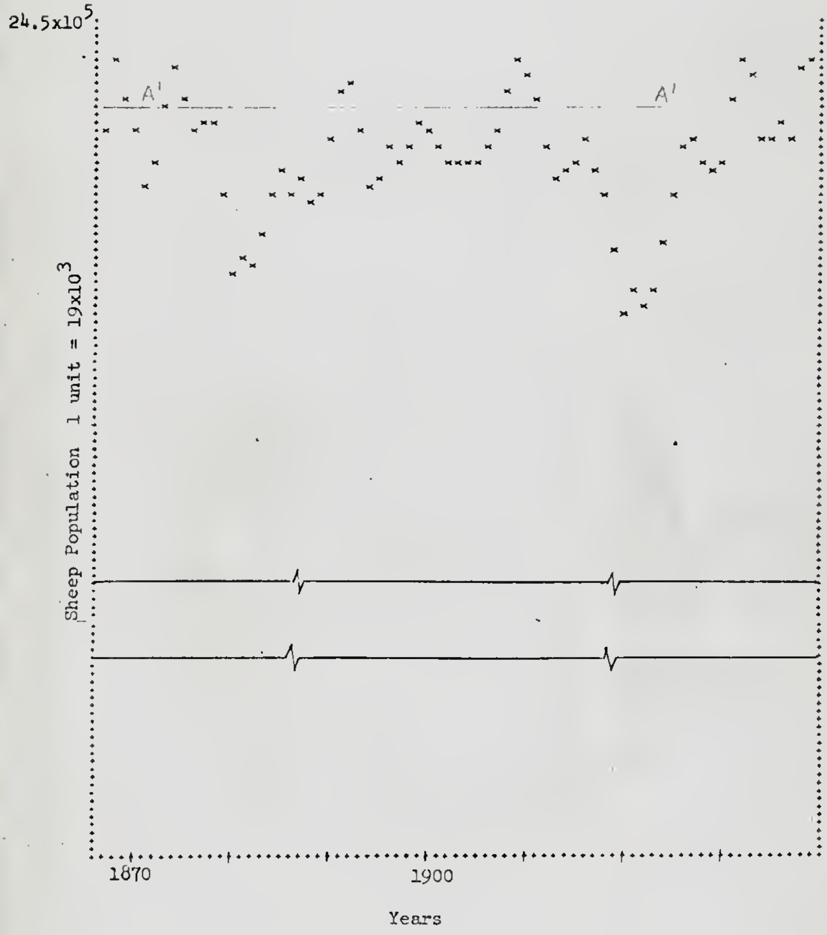

Fig. 4.4 "Detrended" Sheep Data

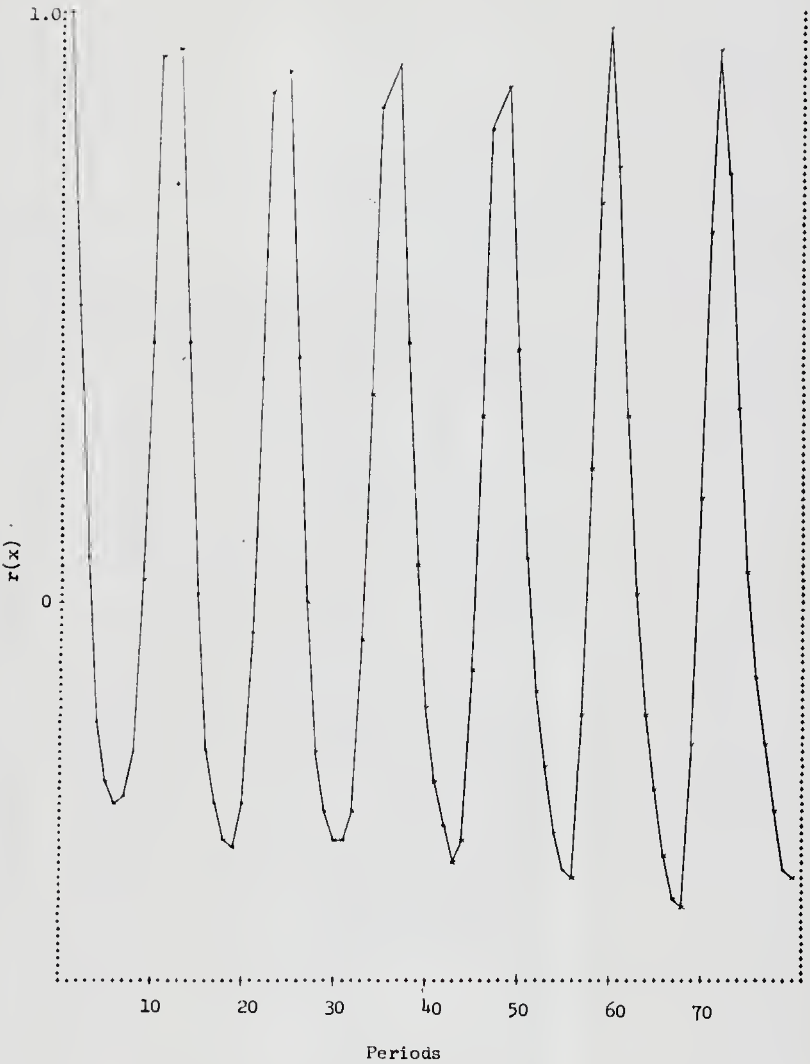

Fig. 4.5 Auto Correlation Function for Airline Data

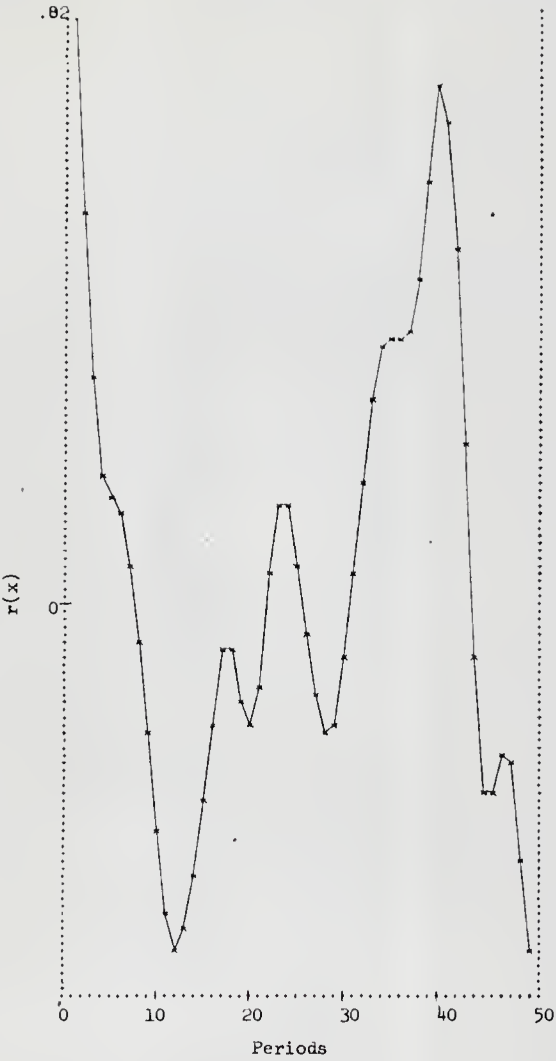

Fig. 4.6 Autocorrelation Function for Sheep Data

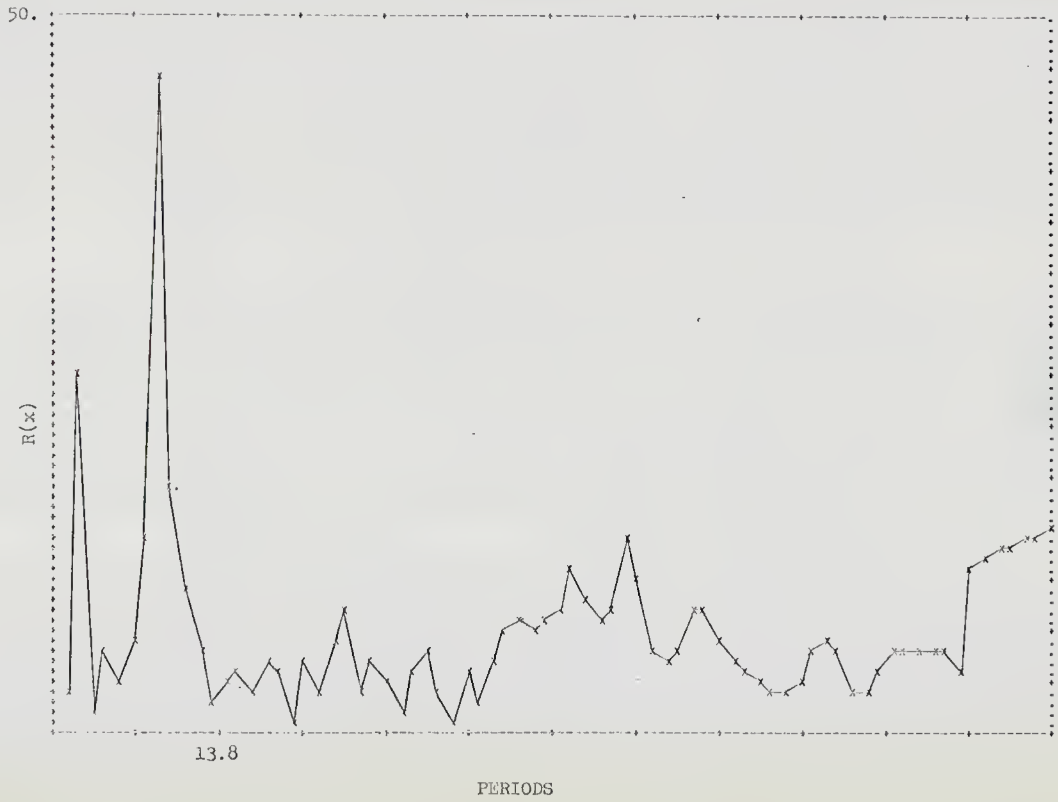

Fig. 4.7 Power Spectrum for International Airline Data

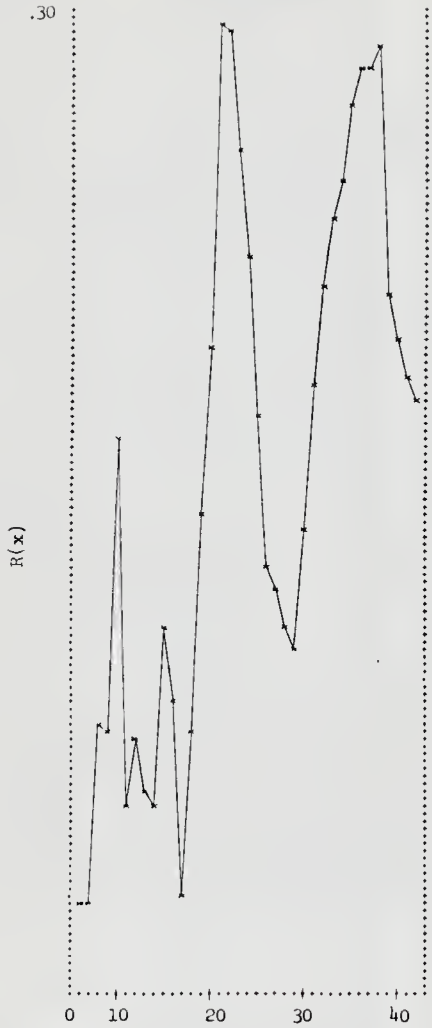

Fig. 4.8 Power Spectrum for Sheep Data

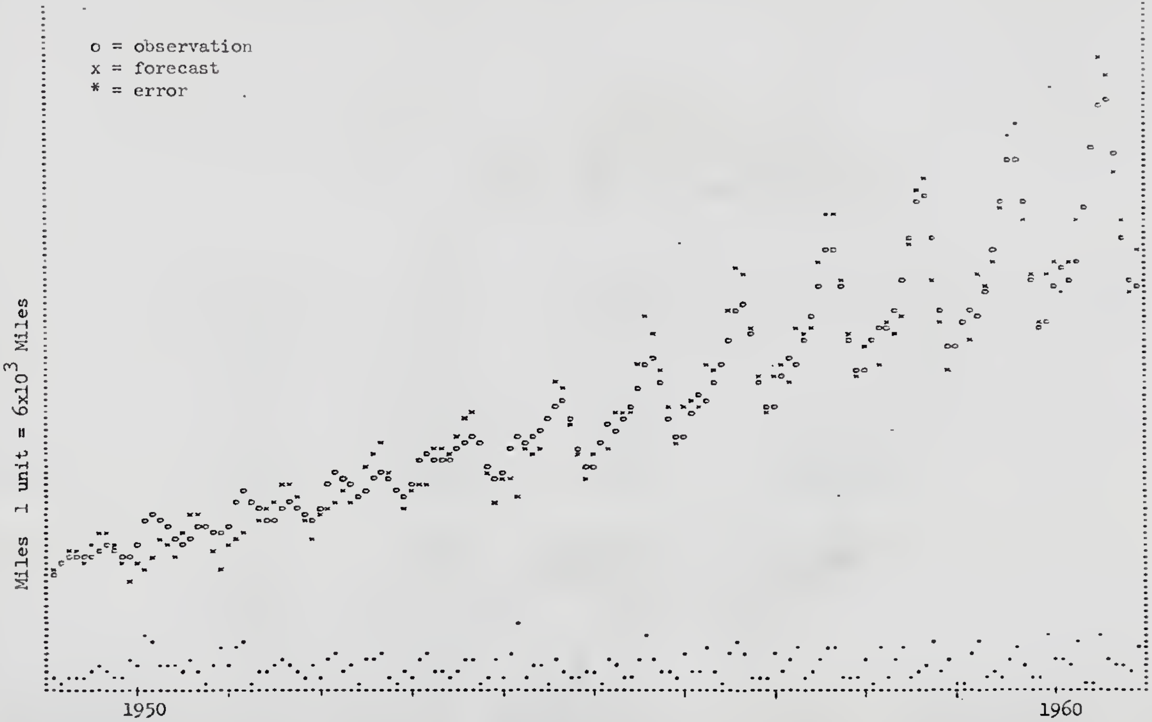

Fig. 4.9 Forecast for International Airline Data BETA = 0.7

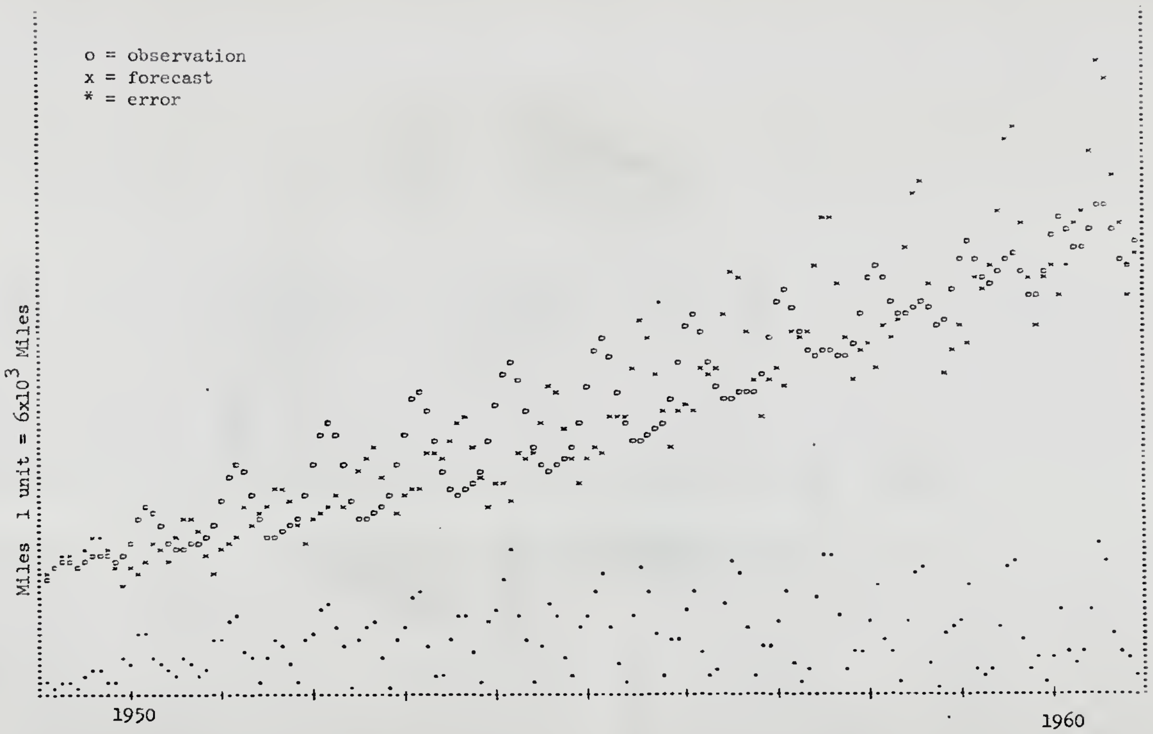

Fig. 4.10 Forecast for International Airline Data BETA = 0.9

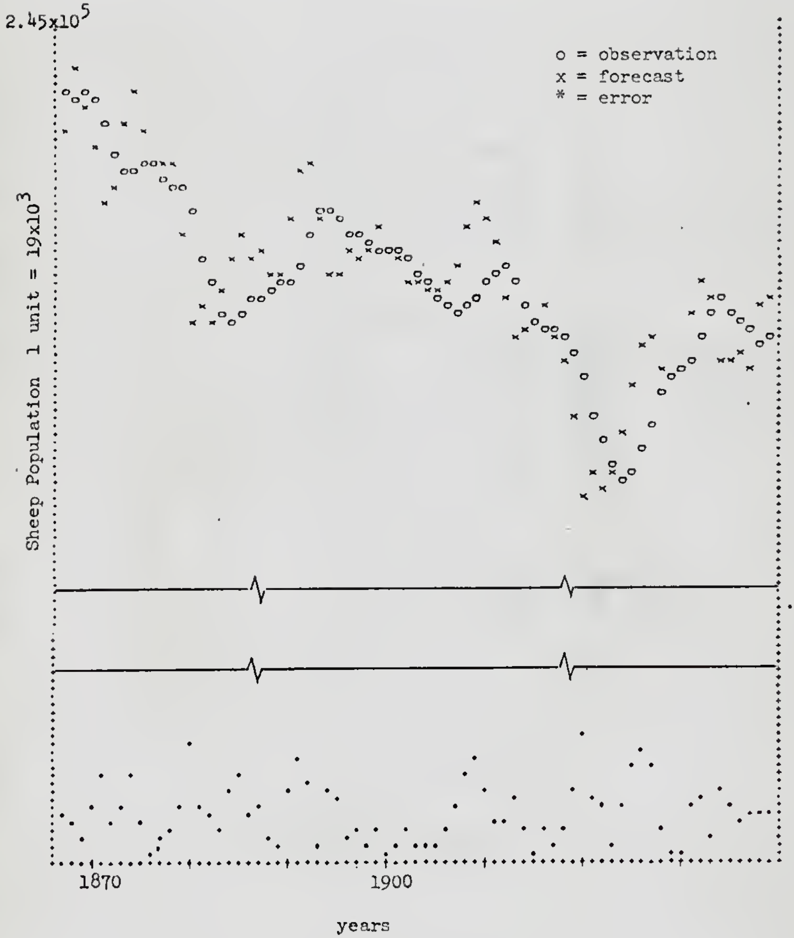

Fig. 4.11 Forecast of Sheep Data

DISCUSSION OF RESULTS

The results of this investigation indicate that exponential smoothing can be used successfully and practicably in forecasting time series. However, the effectiveness of exponential smoothing must be qualified by three conditions

(1) The correct choice of the parameters of the forecasting model,

(2) An adequate choice of the value of the smoothing constant and

(3) An adequate range of the data.

If any of these conditions is not met, exponential smoothing will not yield accurate forecasts.

Brown presents the application of general exponential smoothing as a means for forecasting a time series. This investigation is concerned with developing techniques to aid in the selection of the forecasting model and demonstrating the sensitivity of the forecasts obtained by exponential smoothing to the parameters of the forecast. Trend analysis was developed and it was found that this technique is quite useful in selecting the terms in the model which describe the trend in the mean and the trends in the periodic component. In the time series considered linear trend removal was used. However, general polynomial trend removal was developed and it is expected that this extension can be very useful.

Spectral analysis is another technique used for determining the forecasting model. This technique is compared with the autocorrelation analysis presented by Brown and found to be much more effective for selecting the periodic terms in the forecasting model. The combination of spectral analysis with "detrending" to remove the effect of the trend of the mean is an original contribution of this research and is considered one of the most

important developments. In the case of the sheep data it was shown that an incorrect choice of the basic period of the forecasting model negates the effect of the terms which describe the periodic component of the time series and even makes the inclusion of these terms detrimental to the forecast. When the correct choice of the basic period is obtained by spectral analysis, however, these terms increase the accuracy of the forecasts.

The comparison of the forecasts obtained from the airline data and the sheep data indicates that the range of the available data is of primary importance in the application of general exponential smoothing. Through the use of spectral analysis and autocorrelation analysis it was determined that the sheep data contained barely two basic periods whereas the airline data contained twelve periods. Hence, it was determined that the more accurate results obtained in the case of the airline data were due to the requirement that exponential smoothing must be carried out over a sufficient range of the data to be effective.

The effect of the choice of the smoothing constant on the forecasts can be seen in the case of the airline data. When a value of 0.7 was used the forecasts successfully adjusted to the trends in the data. On the other hand a value of 0.9 for this constant relulted in forecasts which degenerated in accuracy with time. This investigation did not propose a method for choosing the value of the smoothing constant but merely points out that a local optimum for a particular time series and set of fitting functions does not exist and that a study of the choice of this constant merits further investigation.

This investigation did not make any assumption about the normality of the distribution of the forecast errors. If the normality assumption was

made then a statistical test of significance such as the F test could be made on these errors. It was noted that the normality assumption, if made, would have to be preceeded by the assumption that the forecasting model is a true representation of the time series and this assumption was considered invalid.

The results of the programs presented demonstrant that the analysis of the data, trend analysis, autocorrelation analysis and spectral analysis can be successfully and economically carried out. Moreover, the program presented for general exponential smoothing shows that a general program which can accomodate changes in the forecasting model, the basic period of the model, the value of the smoothing constant and the time series can be developed and is a valuable decision aid in industrial and economic situations.

ACKNOWLEDGEMENTS

The author would like to acknowledge Dr. L. Eugene Grosh for his assistance and original ideas which made this work possible.

BIBLIOGRAPHY

(1) Brown, Robert G., Smoothing Forecasting and Prediction of Discrete Time Series, Prentice - Hall International, Inc., New Jersey, 1963.

(2) Carslaw, Horatio S., Introduction to the Theory of Fourier Series, Macmillan and Co., Limited, London, 1930.

(3) Davis, Harold T., The Analysis of Economic Time Series, The Principia Press, Inc., Bloomington, Indiana, 1941.

(4) Davis, Harold T., The Theory of Econometrics, the Principia Press, Inc., Bloomington, Indiana, 1941.

(5) Dixon, Wilfrid J. and Massey, Frank J., Introduction to Statistical Analysis, McGraw - Hill, Inc., New York, 1957.

(6) Hogg, Robert V. and Craig, Allen, T., Introduction to Mathematical Statistics, The Macmillan Company, New York, 1965.

(7) Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley & Sons, Inc., New York, 1954.

APPENDIX A FOURIER ANALYSIS

The following definition of Fourier analysis is given by Harold Thayer Davis (2,61)

> "The problem of Fourier series is that of representing a function, either continuous or with a finite number of finite discontinuities as exhibited by a set of discrete data, by means of a series of fundamental harmonics."

The above statement refers to a harmonic in the form:

$$y = A\cos(2\pi t/T) + B\sin(2\pi t/T)$$

where T is the period of the harmonic. This expression can be written in the form

$$y = A^2 + B^2 \cos(2\pi t/T - \alpha)$$

where α is the lag angle given by

$$\alpha = \arctan B/A$$

An example illustrating the representation of a harmonic is given in Fig. 1A.

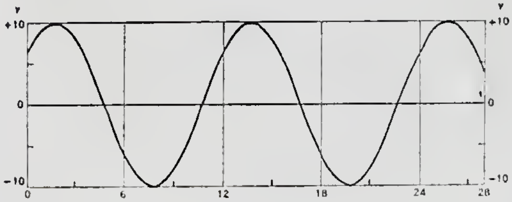

Fig. 1A Representation of a harmonic term

$$y = 6\cos(2\Pi t/12) + 8\sin(2\Pi t/12)$$

$$= \left[\sqrt{(6)^2 + (8)^2}\right] \cos\left[(2\Pi t/12) - \alpha\right]$$

where T is the period of the harmonic,

$1/T$ is the frequency,

$(A^2 + B^2)^{\frac{1}{2}}$ is the amplitude,

α is the phase angle .

The purpose of defining the harmonic term lies in the ability to represent a function by a series of these harmonic terms. The series composed of these harmonics is called the Fourier series. The value of the Fourier series lies in the following basic theorem

> If $f(t)$ is a single-valued function which has a derivative throughout the interval $-a \leqq t \leqq a$ except for a finite number of points at which it has finite discontinuities, and for other values it is defined by the equation
>
> $$f(t) = f(t + 2a)$$
>
> then $f(t)$ can be represented by the series

$$y = \tfrac{1}{2}A_0 + A_1\cos(\Pi t/a) + A_2\cos(2\Pi t/a) + A_3\cos(3\Pi t/a) + \ldots$$
$$+ B_1\sin(\Pi t/a) + B_2\sin(2\Pi t/a) + B_3\sin(3\Pi t/a) + \ldots$$

This series is the Fourier series. The coefficients are determined from the integrals

$$A_n = 1/a \int_{-a}^{a} f(s)\cos(n\Pi s/a)ds \quad ,$$

$$B_n = 1/a \int_{-a}^{a} f(s)\sin(n\Pi s/a)ds \quad .$$

In the case above the continuous series is considered. When analyzing data such as a time series a transformation to the discrete case is in order. Moreover, it is often useful to convert from the symmetric form shown above to another, more suitable form. Since in a time series the data is given over the range

$$0 \leqq t \leqq 2a$$

the transformation

$$g(t) = f(t-a)$$

is made and the integrals are rewritten as

$$A'_n = 1/a \int_0^{2a} g(t) \cos(n\Pi t/a)dt ,$$

$$B'_n = 1/a \int_0^{2a} g(t) \sin(n\Pi t/a)dt .$$

If the data is given in the form

$$f_1, f_2, f_3, \ . \ . \ . \ , f_N \ .$$

Then the Fourier coefficients can be conviently represented in the form given by Davis (2.64).

$$A_n = 2/N \sum_{t=1}^{N} f_t \cos(2\Pi t/N) ,$$

$$B_n = 2/N \sum_{t=1}^{N} f_t \sin(2\Pi t/N) .$$

APPENDIX B REGRESSION ANALYSIS

In the case of linear regression a straight line is fitted to the data in such a way as to minimize the sum of the squares of errors (6). In this case the error is defined as the difference between the actual data point and the value obtained by the straight line at that point.

The equation for the straight line is utilized in the form

$$y' = a + b(x - \bar{x}) \tag{B-1}$$

where b is the slope of the line and a is the Y intercept on the line $x = \bar{x}$. The problem is to determine the parameters a and b so that the sum of the squares of the errors of estimation will be a minimum. Let the co-ordinates of the i^{th} point be denoted by (x_i, y_i). Then the term to be minimized is

$$\sum_{i=1}^{n} (y_i - y_i')^2$$

where y_i' is determined by B-1.

If this function is denoted by G(a,b) and written as

$$G(a,b) = \sum_{i=1}^{n} \left(y_i - a - b(x_i - \bar{x})\right)^2$$

then the conditions for the above expression to be a minimum are that its partial derivatives vanish. Hence a and b must satisfy the equations

$$\frac{\partial G}{\partial a} = \sum 2 \left(y - a - b(x - \bar{x})\right)\left(-1\right) = 0$$

$$\frac{\partial G}{\partial b} = \sum 2 \left(y - a - b(x - \bar{x})\right)\left(-(x - x)\right) = 0$$

When the summations are performed term by term and the sums that involve y are transposed, these equations assume the form

$$an + b\sum (x-\bar{x}) = \sum y$$

$$a\sum (x-\bar{x}) + b\sum (x-\bar{x})^2 = \sum (x-\bar{x})y$$

Since $\sum(x-x) = 0$, the solution of these equations is given by

$$a = \bar{y} \text{ and } b = \frac{\sum(x-\bar{x})y}{\sum(x-\bar{x})^2}$$

For computational purposes, it is convenient to change the form of the expression for b in the following manner

$$b = \frac{\sum xy - \bar{x}\sum y}{\sum x^2 - 2\bar{x}\sum x + \sum \bar{x}^2}$$

$$= \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

The concept of linear regression can be easily extended to polynomial regression. (6) Let the degree of the polynomial be k and let the equation of the polynomial be written in the form

$$Y' = c_0 + c_1X + c_2X^2 + \ldots + c_kX^k$$

As in the case of linear regression the unknown coefficients are estimated by the method of least squares. This is equivalent to minimizing the sum

$$\sum_{i=1}^{n} (Y_i - Y'_i)^2$$

Since it is more convenient to work with variables measured from their sample means than with the variables themselves, the following definitions are made

$$y = Y - \bar{Y}$$

$$x_j = X_j - \bar{X}_j$$

If y' is defined by $y' = Y' - \bar{Y}$, then

$$Y - Y' = y + \bar{Y} - (y' + \bar{Y}) = y - y'$$

If now the capital X's and Y's are expressed in terms of the small x's and y's the polynomial regression model can be written in the form

$$y' = a_0 + a_1 x + a_2 x^2 + . . . + a_k x^k \qquad \text{B-2}$$

Since minimizing $\sum(Y - Y')^2$ is equivalent to minimizing $\sum(y - y')^2$ it is just as well to determine the a's to minimize the latter sum which because of B-2 may be written as

$$G(a_0, a_1, . . . a_k) = \sum\left(y - a_0 - a_1 x - . . . . - a_k x^k\right)^2$$

If this function is to have a minimum value, it is necessary that its partial derivatives vanish there. Hence, the a's must satisfy the equations

$$\frac{\partial G}{\partial a_0} = \frac{\partial G}{\partial a_1} = . . . = \frac{\partial G}{\partial a_k} = 0$$

This differentiation produces the equations

$$\sum 2(y - a_0 - a_1 x \ . \ . \ . \ a_k x^k)(-1) = 0$$

$$\sum 2(y - a_0 - a_1 x - \ . \ . \ . \ . - a_k x^k)(-x) = 0$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

$$\sum 2(y - a_0 - a_1 x - \ . \ . \ . \ . - a_k x^k)(-x^k) = 0$$

If these equations are multiplied by ½, the summations performed term by term, and the first sum transfered to the right side, these equations will assume the form

$$a_0 n + a_1 \sum x + \ . \ . \ . + a_k \sum x^k = \sum y$$

$$a_0 \sum x + a_1 \sum x^2 + \ . \ . \ . + a_k \sum x^{k+1} = \sum xy$$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

$$a_0 \sum x^k + a_1 \sum x^{k+1} + \ . \ . \ . + a_k \sum x^{2k} = \sum x^k y$$

Since

$$\sum x^j = \sum (X^j - \bar{X}) = 0$$

and

$$\sum y = \sum (Y - \bar{Y}) = 0 \ ,$$

all terms in the first equation except the first term vanish. This implies that $a_0 = 0$, and thus the number of equations to be solved has been reduced by one. The problem is now reduced to solving the equations

$$a_1\sum x^2 + a_2\sum x^3 + \ldots + a_k\sum x^{k+1} = \sum xy$$

$$a_1\sum x^3 + a_2\sum x^4 + \ldots + a_k\sum x^{k+2} = \sum x^2 y$$

$$\ldots\ldots\ldots\ldots$$

$$a_1\sum x^{k+1} + a_2\sum x^{k+2} + \ldots + a_k\sum x^{2k} = \sum x^k y$$

Having developed the expressions for the coefficients in the regression model for the general polynomial, the next step is to consider a more general regression model. Suppose the model contains terms of the following types

(1) polynomials

(2) trigonometric functions

(3) exponential functions

(4) emperical functions

In this case the vector of coefficients can be represented as

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix}$$

and the fitting functions are represented in vector notation as

$$f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \cdot \\ \cdot \\ \cdot \\ f_m(t) \end{bmatrix}$$

where $f_i(t)$ is the value of the i^{th} fitting function evaluated at time t. Hence, the data is represented by the model

$$\hat{x}_t = \sum_{i=1}^{m} a_i f_i(t) = a'f(t)$$

The criterion for the selection of the coefficients is taken as

$$\min \sum_{t=1}^{T} w_t^2 e_t^2$$

where e(t) is the residual defined as

$$e(t) = x(t) - a'f(t)$$

and w_t is the weight given the residual at time t.

The above expressions can easily be put into matrix form. A matrix $\mathcal{F}$ is defined as an m x T matrix of elements $f_i(t)$, the value of the i^{th} fitting function at time t. A row vector $\hat{x}$ is defined to be the sequence $(x_1, x_2, \ldots, x_T)$ of values given by the model a' . Moreover, e is defined as the sequence $(e_1, e_2, \ldots, e_T)$ of residuals, where

$$e_t = x_t - \hat{x}_t = x_t - a'(T)f(t)$$

In order to find the expression for the coefficient vector that satisfies the regression criteria matrix notation is continued. Let W be a TxT matrix in which W_{ii} is the square root of the weight given the residual at time i. All off diagonal elements of W are zero. Noting that the expression of the model is

$$a'\mathcal{F} = x - e$$

the residuals can be expressed as

$$e = x - a'\mathcal{Z}.$$

Now for a particular choice of a the sum of squares is

$$Sa = ee' = (xW - a'\mathcal{Z}W)(xW - a'\mathcal{Z}W)'$$

$$= x^2\ W'W - 2a'x\mathcal{Z}WW' + (a')^2\ (\mathcal{Z}W)(\mathcal{Z}W')$$

The particular set of coefficients that minimize this sum is found from

$$\frac{\partial Sa}{\partial a} = -xWW'\mathcal{Z} + a'\mathcal{Z}W(\mathcal{Z}W)' = 0$$

Hence

$$xWW'\mathcal{Z}' = a'\mathcal{Z}WW'\mathcal{Z}'$$

By denoting the nxn symmetric matrix as

$$\mathcal{Z}W(\mathcal{Z}W)' = \sum_{t=1}^{T} w_t^2\ f(t)f'(t) = F$$

then

$$xWW'\mathcal{Z}' = a'F \qquad \text{B-3}$$

Now the conditions for F to have an inverse are

(1) There are at least as many observations, x, as degrees of freedom in the model

(2) The fitting functions are linearly independent.

If the above two conditions hold true then by postmultiplication of B-3 by F^{-1}

$$a' = xW^2\mathcal{Z}'F^{-1}$$

This is the expression for the vector of coefficients that minimize the weighted sum of squared residuals.

APPENDIX C Correlation and Autocorrelation

Given the two variables

$$x_i = x_1, x_2, \ldots, x_n$$

$$y_i = y_1, y_2, \ldots, y_n$$

The correlation coefficient r may be expressed as

$$r = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \;/\; (n)S_x S_y$$

$$\text{where } S_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \;/\; n , \qquad \text{C-1}$$

$$S_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \;/\; n .$$

If the variables are first converted to standard form then C-1 may be expressed more conveniently. That is if

$$u_i = (x_i - \bar{x}) \;/\; S_x$$

$$v_i = (y_i - \bar{y}) \;/\; S_y$$

then

$$r = \sum_{i=1}^{n} u_i v_i / n \qquad \text{C-2}$$

Two properties of the correlation coefficient are significant

(1) $-1 \leq r \leq +1$

(2) $r = \pm 1$ only if the points lie on a straight line.

Hence, the correlation coefficient is a measure of the strength of the linear relationship between the two variables. If the correlation coefficient is 1 then the two variables are directly linearly related. If it is -1 then an inverse linear relationship exists. A correlation coefficient

of zero indicates that no linear relationship exists between the two variables.

The idea of correlation between two variables can be extended to the consideration of the lag correlation of a single variable. Consider the variable $x(t)$, $t = 1,2, \ldots, n$, where $x(t)$ represents the value of x at time t. The idea of autocorrelation is a measure of the correlation, or the strength of the linear relationship, between $x(t)$ and $x(t+k)$. Consider the expression

$$\sum_{j=k+1}^{T} x(j)x(j-k)$$

where T is the range of the series. The expected value of this term is called the average lagged product. Moreover, if the series has been adjusted so that the expected value, $E(\hat{x})$, is zero, where $\hat{x}$ denotes the adjusted series, then the average lagged product is the autocovariance

$$R_{xx}(k) = \sum_{j=k+1}^{T} \hat{x}_j \hat{x}_{j-k} \ / \ T-1-k$$

where $R_{xx}(k)$ is the autocovariance evaluated for lag k.

Since the variance of a sequence of numbers which have been adjusted to have an average value of zero is just the expected value of their squares, $E(\hat{x}^2)$, by virtue of the notation just defined this can be expressed as $R_{xx}(0)$. Hence, by C-2 the autocorrelation coefficient can be expressed as

$$\rho(k) = R_{xx}(k)/R_{xx}(0) \ .$$

The set of values of this function for all lags $k = \pm 1, \pm 2, \ldots$ is called the autocorrelation coefficient.

APPENDIX D Exponential Smoothing

The smoothed statistic for exponential smoothing is defined as

$$S_t(x) = \alpha x(t) + (1-\alpha)S_{t-1}(x)$$

where $x(t)$ is the observation at time t $\quad 0 \leq \alpha \leq 1$

α is an undimensioned positive ratio less than 1.

If the above definition of exponential smoothing is defined as single smoothing then double smoothing can be defined as

$$S_t^{(2)}(x) = \alpha S_t^{(1)}(x) + (1-\alpha)S_{t-1}^{(2)}(x) .$$

Similiarly multiple smoothing of order k is defined as

$$S_t^{(k)}(x) = S_t^{(k-1)}(x) + (1-\alpha)S_{t-1}^{(k)}(x)$$

The fundamental theorem of exponential smoothing proves that it is possible to estimate the (n+1) coefficients (derivatives) in an n^{th} order polynomial model by linear combinations of the first n+1 smoothed statistics.

THE FUNDAMENTAL THEOREM OF EXPONENTIAL SMOOTHING

If the observations $x(t+\tau)$ are represented by the model (1,133)

$$x(t+\tau) = \sum_{k=0}^{n} \tau^k \, x_t^{(k)}/k!$$

where τ is the forecast period then

$$S_t^{(p)}(x) = \sum_{k=0}^{n} (-1)^k \, (x_t^{(k)}/k!) \, \alpha p/(p-1)! \, \sum_{j=0}^{\infty} j^k \beta^j (p-1+j)!/j! \qquad \text{D-1}$$

where $x_t^{(k)}$ is the k^{th} derivative evaluated at time t.

As a proof to the fundamental theorem of exponential smoothing two vectors are defined, a vector x which represents the infinite sequence of observations $x(t)$ for $t = -,\ldots -1,0,1,\ldots,\infty$, and a vector S with components

$$S_t = \begin{cases} 0 & (t < 0) \\ \alpha\beta^t & (t \geq 0) \end{cases} .$$

Exponential smoothing can be represented as a convolution of the two vectors $x * S$ which has the components

$$(x * S)_t = S_t(x) = \sum_{j=0}^{\infty} x(t-j)S_j = \alpha \sum_{j=0}^{\infty} \beta^j x(t-j)$$

Having defined the convolution relationship for single smoothing, since the convolution operation is associative, multiple smoothing of order p is equivalent to the convolution $x * S^{(p)}$ where S^p necessarily has the components

$$(S^{(p)})_t = \begin{cases} 0 & (t < 0) \\ \alpha^p\beta^t\ (p-1+t)!/t!(p-1) & (t \geq 0) \end{cases}$$

Therefore

$$S_t{}^p(x) = \sum_{j=0}^{\infty} x(t-j)\ (\alpha^p\beta^j(p-1+j)!)/j(p-1)!\ .$$

But since

$$x(t-j) = \sum_{k=0}^{n} (-1)(x_t^{(k)}/k!)\ j^k$$

the theorem is proved.

The first five smoothed statistics can be written as

$$S_t(x) = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \alpha \sum_{j=0}^{\infty} j^k \beta^j$$

$$S_t^{(2)}(x) = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \alpha^2 \sum_{j=0}^{\infty} j^k (j+1) \beta^j$$

$$S_t^{(3)}(x) = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \frac{\alpha^3}{2} \sum_{j=0}^{\infty} j^k (j+1)(j+2) \beta^j$$

$$S_t^{(4)}(x) = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \frac{\alpha^4}{6} \sum_{j=0}^{\infty} j^k (j+1)(j+2)(j+3) \beta^j$$

$$S_t^{(5)} = \sum_{k=0}^{n} (-1)^k \frac{x_t^{(k)}}{k!} \frac{\alpha^5}{24} \sum_{j=0}^{\infty} j^k (j+1)(j+2)(j+3)(j+4) \beta^j$$

From the structure of these terms the following is determined.

(1) The p^{th} order of smoothing is given as the alternating sum of the n coefficients in the Taylor series

$$x_t^{(k)}/k!$$

(2) The coefficients of these derivatives are infinite sums involving the smoothing constant.

By referring to the closed form of the infinite sums (1,135).

$\underline{n}$	Form	Sum
0	$\sum \beta^j$	$1/1-\beta$
1	$\sum j\beta^j$	$\beta/(1-\beta)^2$
2	$\sum j^2\beta^j$	$\beta(1-\beta)/(1-\beta)^3$
3	$\sum j^3\beta^j$	$\beta(1+4\beta+\beta^2)/(1+\beta)^4$
4	$\sum j^4\beta^j$	$\beta(1+11\beta+11\beta^2+\beta^3)/(1+\beta)^5$
5	$\sum j^5\beta^j$	$\beta(1+26\beta+66\beta^2+26\beta^3+\beta^4)/(1+\beta)^6$
6	$\sum j^6\beta^j$	$\beta(1+57\beta+302\beta^2+302\beta^3+57\beta^4+\beta^5)/(1+\beta)^7$

and writing the vector of coefficients as

$$a = \begin{bmatrix} a_0^{(t)} \\ a_1^{(t)} \\ \\ \cdot \\ \cdot \\ \cdot \\ a_n^{(t)} \end{bmatrix} = \begin{bmatrix} x_t^{(0)}/0! \\ x_t^{(1)}/1! \\ \\ \cdot \\ \cdot \\ \cdot \\ x_t^{(n)}/n! \end{bmatrix}$$

the fundamental theorem can be expressed as

$$S_t = Ma \qquad \text{D-2}$$

where M is an nxp matrix with elements involving infinite sums of powers of the smoothing constant

$$M_{ik} = \frac{\alpha^i}{(i-1)!} \sum_{j=0}^{\infty} j^k \beta^j \; \frac{(i-1+j)!}{j!}$$

where M_{ik} is the element in the i^{th} row and k^{th} column of the matrix M

S_t is the vector of the p smoothed statistics evaluated at time t. By necessity n=p.

a is the vector of the coefficients

With the fundamental theorem in matrix form the n+1 simultaneous equations in the n+1 derivatives given in D-1 can be solved for these derivatives in terms of the n+1 smoothed statistics. If D-2 is post-multiplied by M^{-1} the result is

$$a = S_t M^{-1}.$$

APPENDIX E

The purpose of this appendix is to present the computer programs. The programs are referred to by figure numbers in the following manner.

E-1	The first phase of the program for general exponential smoothing - "INCONT"
E-2	A model for a linear trend plus a superimposed sinusoid (included in Phase I)
E-3	Phase II - "RAY"
E-4	Phase III - "MATINV"
E-5	The subprogram to calculate the h vector - "HVEC"
E-6	Phase IV - "FORCST"
E-7	Phase V - "PLOTTER"
E-8	The subprogram to plot the observations, forecasts and errors - "PLOTS"
E-9	Monitor cards used for phasing in the PR-155 system
E-10	The program for calculation the autocorrelation function
E-11	The subprogram to remove the trend from the time series - "TREND"
E-12	The program to calculate the power spectrum

```
    MON$$      JOB
    MONS$      ASGN MJB,12
    MONS$      ASGN MGO,16
    MONS$      MODE GO,TEST
    MONS$      EXEQ FORTRAN,,,09,,,,INCONT
      DIMENSIONCHK(10)
      DIMENSIONX(500)
      DIMENSIONCI(9),A(9,1),TM(9,9)
  910 FORMAT(1X,10F10.5)
  999 FORMAT(1X,6I10)
  998 FORMAT(1X,F30.7)
  929 FORMAT(F5.1)
  447 FORMAT(50X,I4)
    1 FORMAT(I4)
C CHOOSE THE NEXT MODEL TO BE TRIED
      REWIND7
      NNN=1
C READ IN THE OBSERVATIONS
      READ(1,1)(ND)
      DO414I=1,ND
  414 READ(1,929)(X(I))
C READ IN THE PERIOD NTOU
      READ(1,1)(NTOU)
  625 IF(NNN-2)77,88,88
   88 STOP
   77 GO TO(33,66),NNN
```

FIG. E-1

```
   33 CONTINUE
      MN=4
      N=4
      CI(1)=2297.
      CI(2)=-12.2
      CI(3)=0.0
      CI(4)=0.0
      DO3I=1,NTOU
      TOU=NTOU
      B=I
      C3=X(I)*SIN(2.*3.14*B/TOU)
      CI(3)=CI(3)+C3
      C4=X(I)*COS(2.*3.14*B/TOU)
    3 CI(4)=CI(4)+C4
      CI(3)=(2./TOU)*CI(3)
      CI(4)=(2./TOU)*CI(4)
      A(1,1)=1.
      A(2,1)=0.
      A(3,1)=0.
      A(4,1)=1.
      CHK(1)=0.0
      CHK(2)=1.0
      CHK(3)=1.0
      CHK(4)=0.0
C     READ IN THE TRANSITION MATRIX
      TM(1,1)=1.0
      TM(1,2)=0.0
      TM(1,3)=0.0
      TM(1,4)=0.0
      TM(2,1)=1.0
      TM(2,2)=1.0
      TM(2,3)=0.0
      TM(2,4)=0.0
      TM(3,1)=0.0
      TM(3,2)=0.0
      TM(3,3)=COS(2.*3.14/TOU)
      TM(3,4)=SIN(2.*3.14/TOU)
      TM(4,1)=0.0
      TM(4,2)=0.0
      TM(4,3)=-SIN(2.*3.14/TOU)
      TM(4,4)=COS(2.*3.14/TOU)
      GOTO818
   66 CONTINUE
  818 WRITE(7)ND,NTOU,N,(A(I,1),I=1,N)
      ,(X(I),I=1,ND),(CI(I),I=1,N),NNN,
     1MN,(CHK(I),I=1,N)
      DO817I=1,N
  817 WRITE(7)(TM(I,J),J=1,N)
      NNN=NNN+1
      GO TO 625
      END
```

FIG. F-2

```
      MON$$      EXEQ FORTRAN,..09....RAY
        INTEGERRNP
        DIMENSIONAA(10,1)
        DIMENSIONCHK(10)
        DIMENSIONEK(9,9),C11(9,9)
        DIMENSIONA(9,1),P(1,9),C1(9,9),E(9,9)
        DIMENSIONCOI(9)
        DIMENSIONX(500),CI(9),A1(9,1),H(9,1)
        DIMENSIONAO(9,1),TM(9,9),AF(9,1)
      1 FORMAT(I4)
    111 FORMAT(1X,10F11.4)
    112 FORMAT(1X,I4)
    999 FORMAT(1X,6I10)
    998 FORMAT(1X,F30.7)
    910 FORMAT(1X,10F10.5)
    662 FORMAT(40X,I2)
    444 FORMAT(5F15.5)
    625 FORMAT(1X,12F9.5)
     25 FORMAT(30X,F20.4)
     23 FORMAT(5F20.4)
     24 FORMAT(10X,5F20.4)
    624 FORMAT(1X,18H E INVERSE IS      )
        REWIND7
        REWIND5
        REWIND6
        NPN=1
        RNP=0
C*****A NEW MODEL IS READ IN STARTING HERE.
C*****BETA IS INITIZLIZED TO 0.7
     46 ABETA=0.70
        READ (7)ND,NTOU,N,(A(I,1),I=1,N)
       1,(X(I),I=1,ND),(CI(I),I=1,N),NNN,
       1MN,(CHK(I),I=1,N)
        DO817I=1,N
    817 READ (7)(TM(I,J),J=1,N)
        TN=N
        AP=(1./TN)
C   SAVE THE INITIAL VALUES OF THE CONSTANTS
        DO934I=1,N
    934 COI(I)=CI(I)
C*****SAVE THE INITIAL VALUES
C*****OF THE FITTING FUNCTIONS
        DO77I=1,N
     77 AO(I,1)=A(I,1)
C*****STARTING HERE THE E MATRIX IS
C*****RECALCULATED WITH A NEW BETA
     67 OBETA=(ABETA)**AP
C*****REINITIALIZE THE VECTOR
C*****OF FITTING FUNCTIONS
        DO42I=1,N
        CI(I)=COI(I)
     42 AA(I,1)=AO(I,1)
```

FIG. E-3

```
C*****COMPUTE THE F MATRIX FOR
C*****ONE MODEL AND ONE BETA
      DO8I=1,N
      DO8J=1,N
      FK(I,J)=0.0
    8 F(I,J)=0.0
      BETA=1.0
      CNT=1.
      K=1
   99 DO30I=1,N
   30 B(1,I)=A(I,1)
      DO40I=1,N
      M=I
      DO40J=M,N
      C1(I,J)=A(I,1)*B(1,J)*(BETA)
   40 CONTINUE
      DO43I=1,N
      M=I
      DO43J=M,N
   43 C1(I,J)=C1(I,J)*BETA
      IF(K-1)66,16,66
C*****CHECK THE CONVERGENCE OF THE
C*****F MATRIX AT EACH 50TH ITERATION.
C*****THIS ROUTINE CHECKS UNTIL IT
C*****FINDS AN ELEMENT WHICH IS
C*****NOT STABILIZED
   66 IF(CNT-50.)16,44,16
   44 DO80I=1,N
      CNT=0.
      M=I
      DO80J=M,N
      DENM=F(I,J)
      ANUM=C1(I,J)
      RATIO=ANUM/DENM
      IF(RATIO-.000001)80,80,16
   80 CONTINUE
C*****NORMAL EXIT FROM THIS LOOP OCCURS
C*****WHEN ALL THE ELEMENTS OF THE F MA-
C*****TRIX MEET THE CONVERGENCE CRITERIA.
C*****WHEN THIS OCCURS THE F MATRIX IS
C*****COMPLETED BY SYMMETRY AND THE PROGRAM
C*****CHECKS TO SEE OF THE RANGE OF BETA
C*****HAS BEEN COVERED
      DO87I=2,N
      L=I-1
      DO87J=1,L
      FK(I,J)=FK(J,I)
   87 F(I,J)=F(J,I)
      WRITE(6)NNN,M,ND,( C1(I),I=1,N),
     1,(A1(I,1),I=1,N),(X,I),I=1,ND),,ABETA,YN
      DO69I=1,N
   69 WRITE(6)(TY(I,J),J=1,N)
```

FIG. F-3 (CONTINUED)

```
      WRITE(5)N,(AO(I,1),I=1,N),OBETA
      DO68I=1,N
   68 WRITE(5)(F(I,J),J=1,N)
      DO48I=1,N
   48 WRITE(5)(FK(I,J),J=1,N)
C*****PARAMETRIC ON BETA
      IF(ABETA-0.70)102,45,45
   45 RNR=RNR+1
      IF(RNR-NRN)46,14,14
   14 STOP
  102 ABETA=ABETA+0.20
C*****REPLENISH THE VECTOR OF
C*****FITTING FUNCTIONS AND USE A NEW
C*****BETA
      GO TO 67
C*****THE PROGRAM BRANCHES TO HERE WHEN
C*****THE CONVERGENCE CRITERIA IS NOT MET
C*****AND NORMAL ITERATION CONTINUES.
   16 DO50I=1,N
      M=I
      DO50J=M,N
      FK(I,J)=FK(I,J)+C11(I,J)
   50 F(I,J)=F(I,J)+C1(I,J)
C*****UPDATE THE VECTOR OF FITTING FUN-
C*****CTIONS USING  THE TRANSITION MA-
C*****TRIX
      DO31I=1,N
   31 AF(I,1)=0.0
      DO90I=1,N
      DO90J=1,N
   90 AF(I,1)=TM(I,J)*AA(J,1)+AF(I,1)
      DO100I=1,N
      DO100J=1,N
      AA(I,1)=AF(I,1)
  100 A(I,1)=AF(I,1)
C*****THIS ROUTINE CONVERTS THE VECTOR
C*****OF FITTING FUNCTIONS EVALUATED FOR
C*****A FIXED TIME ORIGIN TO THE VECTOR
C*****EVALUATED FOR A MOVING TIME ORIGIN
      DO63I=1,N
      CK=CHK(I)
      IF(CK-1.)63,55,63
   55 A(I,1)=-A(I,1)
   63 CONTINUE
C SAVE F(1) FOR FORECAST
      IF(K-1)62,22,62
   22 DO62I=1,N
      A1(I,1)=AA(I,1)
   62 CONTINUE
```

FIG. F-3 (CONTINUED)

```
BETA=CBETA*BETA
CNT=CNT+1.
K=K+1
GO TO 99
STOP
END
```

FIG. E-3 (CONTINUED)

```
MON$9      EXEQ FORTRAN...09....MATINV
      INTEGERRNP
      DIMENSIONFK(9.9)
      DIMENSIONV2(10.10)
      DIMENSIONV1(10.10)
      DIMENSIONAO(9.1),H(9.1)
      DIMENSIONC(9.9).V(9).F(9.9)
  727 FORMAT(///1X.16H THE H VECTOR IS     )
  789 FORMAT(F20.6)
  444 FORMAT(9F13.3)
  777 FORMAT(///10X.39H THE VARIANCE OF
     1THE COEFFICIENTS IS        )
  888 FORMAT(///1X,11H F INVERSE     )
  889 FORMAT(///20X.6H BETA=.F9.6)
      RNR=0
      NRN=1
      REWIND7
      REWIND 5
   66 READ(5)N.(AO(I.1).I=1.N).OBETA
      DO67I=1,N
   67 READ(5)(F(I,J).J=1.N)
      DO33I=1,N
   33 READ(5)(FK(I,J),J=1,N)
      DO6I=1,N
      DO6J=1.N
    6 C(I,J)=F(I.J)
      NM1=N-1
      DO 6660 L=1,N
      ALL=C(1,1)
      DO 4440 J=1.NM1
 4440 V(J)=C(1.J+1)/ALL
      V(N)=1./ALL
      DO 9990 I=1,NM1
      IP1=I+1
      CIP11=C(IP1,1)
      DO 5550 J=1,NM1
 5550 C(I,J)=C(IP1,J+1) - CIP11*V(J)
 9990 C(I,N)=(-1.0)*CIP11*V(N)
      DO 6660 J=1,N
 6660 C(N,J)=V(J)
```

FIG. F-4

```
      DO16I=1,N
      DO16J=1,N
   16 F(I,J)=C(I,J)
C CALCULATE THE H VECTOR
      CALL HVEC(F,AC,N,H)
C*****WRITE THE EFFECTIVE BETA
      WRITE(3,889)OBETA
C*****WRITE F INVERSE
      WRITE(3,888)
      DO333I=1,N
  333 WRITE(3,444)(F(I,J),J=1,N)
C WRITE OUT THE H VECTOR
      WRITE(3,727)
      DO626I=1,N
      WRITE(3,789)(H(I,1))
  626 WRITE(2,789)(H(I,1))
      DO7I=1,N
    7 WRITE(7)(F(I,J),J=1,N)
      WRITE(7)(H(I,1),I=1,N)
C*****FORM THE V MATRIX
      DO14I=1,N
      DO14J=1,N
   14 V2(I,J)=0.0
      DO11I=1,N
      DO11K=1,N
      DO11J=1,N
   11 V2(K,I)=F(K,J)*FK(J,I)+V2(K,I)
      DO13I=1,N
      DO13J=1,N
   13 V1(I,J)=0.0
      DO12I=1,N
      DO12K=1,N
      DO12J=1,N
   12 V1(K,I)=V2(K,J)*F(J,I)+V1(K,I)
      WRITE(3,777)
      DO122K=1,N
  122 WRITE(3,789)V1(K,K)
      RNR=RNR+1
      IF(RNR-NRN)66,68,68
   68 STOP
      END
```

FIG. E-4 (CONTINUED)

```
  MON$$      EXEQ FORTRAN...GO....HVEC
      SUBROUTINE HVEC(F.AC.N.H)
      DIMENSIONF(9.9).AC(9.1).H(9.1).HF(9.1)
  789 FORMAT(F13.5)
      DO32I=1.N
   32 H(I.1)=0.0
      DO543I=1,N
      DO543J=1,N
      HF(I,1)=F(I,J)*AC(J.1)
  543 H(I,1)=H(I,1)+HF(I.1)
  626 CONTINUE
      RETURN
      END
```

FIG. E-5

```
   MONS$       EXEQ FORTRAN...C9....FORCST
      INTEGERRNR
      DIMENSIONF(9.9)
      DIMENSIONCCI(9)
      DIMENSIONX(500).H(9.1).CI(9).
     1.TM(9.9).CIF(9).CIL(9)
      DIMENSIONA1(9,1),A(9,1)
      DIMENSIONTTM(9.9)
  222 FORMAT(///1X.31H THE VARIANCE OF
     1THE FORECASTS=.F15.6/1H1)
  107 FORMAT(1X.F20.7.F20.4)
  108 FORMAT(1X.15H NEW FORECAST    )
  101 FORMAT(1X.11H MODEL NO.=.I2.20X.6H
     1 BETA=.F4.3)
  198 FORMAT(1X,20H***THE ERROR IS***       )
  973 FORMAT(72H  PERIOD          OBSERVATION
     1    FORECAST       ERROR     CUM ERROR )
  673 FORMAT(F23.5)
  105 FORMAT(10F10.3)
  104 FORMAT(I8,F15.3.F15.2,F15.2.F15.2)
  106 FORMAT(24H SUM OF SQUARE OF ERROR=.F15.3)
      REWIND6
      REWIND7
      REWIND5
      NRN=1
      RNR=0
   66 READ (6)NNN,N.ND,( CI(I).I=1.N),(A1(I.1).I=1.N).
     1(X(I).I=1.ND),ABETA.MN
      WRITE(5)ND,(X(I),I=1.ND)
      WRITE(2.108)
C*****WRITE OUT THE MODEL NUMBER AND
C*****VALUE OF BETA FOR THIS F MATRIX
      WRITE(3,101)MN,ABETA
      DO7I=1.N
    7 READ(6)(TM(I,J).J=1.N)
      DO6I=1.N
    6 READ(7)(F(I,J).J=1.N)
      READ(7)(H(I.1).I=1.N)
C*****MAKE THE TRANSPOSE OF THE TRANSITION
C*****MATRIX
```

FIG. F-6

```
      SUM=0.
      SUMA=0.0
      SUMX=0.0
      DO429I=1,N
      DO429J=1,N
  429 TTM(I,J)=TM(J,I)
C*****MAKE THE FORECAST
      WRITE(3,973)
      DO191K=1,ND
      Z=0.0
      DO711I=1,N
      ZN=CI(I)*A1(I,1)
  711 Z=Z+ZN
      WRITE(5)Z
C*****CALCULATE THE ERROR
      ER=X(K)-Z
C*****CALCULATE THE CUMULATIVE ERROR
      SUM=SUM+ER*ER
      SUMA=SUMA+ER
C*****CALCULATE THE SUM OF THE OBSERVATIONS
      SUMX=SUMX+X(K)
C UPDATE THE CONSTANTS
      DO449I=1,N
  449 CIL(I)=0.0
      DO944I=1,N
      DO944J=1,N
      CIE(I)=TTM(I,J)*CI(J)
  944 CIL(I)=CIL(I)+CIE(I)
      DO926I=1,N
  926 CI(I)=CIL(I)+H(I,1)*ER
      WRITE(3,104)(K,X(K),Z,ER,SUMA)
      WRITE(2,104)(K,X(K),Z,ER)
  191 CONTINUE
      WRITE(3,198)
      WRITE(3,106)(SUM)
      WRITE(2,107)SUM,BETA
C*****CALCULATE THE VARIANCE OF THE FORECASTS
      VAR=SUM/SUMX
      WRITE(3,222)VAR
      RNR=RNR+1
      IF(RNR-NRN)66,68,68
   68 STOP
      END
```

FIG. E-6 (CONTINUED)

```
 MON$$      EXEQ FORTRAN...09....PLOTTER
      INTEGERRNR
      DIMENSIONX(500),FCST(500)
      REWIND5
      RNR=0
      NRN=1
   66 READ(5)ND,(X(I),I=1,ND)
      DO88K=1,ND
      READ(5)Z
   88 FCST(K)=Z
      CALL PLOTS(X,FCST)
      RNR=RNR+1
      IF(RNR-NRN)66,68,68
   68 STOP
      END
```

FIG. F-7

```
MONSS       EXEQ FORTRAN...GO....PLOTS
  SUBROUTINEPLOTS(Y,Z)
  DIMENSIONY(1),Z(1),MP(132)
2 FORMAT(1X,132A1)
1 FORMAT(5A1,3I4)
4 FORMAT(1X,132(1H+))
3 FORMAT(1H1)
5 FORMAT(//10X,40H     TIME SERIES,
  FORECAST,ERROR)
6 FORMAT(//10X,13HX=TIME SERIES,10X,
  10HO=FORECAST, 10X,7H*=ERROR)
7 FORMAT(//10X,4HMAX=,I4,5X,4HMIN=,
  I3,5X,6HSCALE=,I3)
  WRITE(3,3)
  READ(1,1)MS1,MS2,MS3,MA,MB,MAX,MIN,NPTS
  WRITE(3,4)
  NSCAL=(MAX-MIN)/130
  NSAL=MIN-NSCAL*5/2
  MP(1)=MA
  DO11I=1,NPTS
  J=(IFIX(Y(I))-NSAL)/NSCAL
  K=(IFIX(Z(I))-NSAL)/NSCAL
  M=IABS(J-K)+2
  MP(J)=MS1
  MP(M)=MS3
  MP(K)=MS2
  WRITE(3,2)(MP(L),L=1,132)
  MP(M)=MB
  MP(J)=MB
11 MP(K)=MB
  WRITE(3,4)
  WRITE(3,5)
  WRITE(3,6)
  WRITE(3,7)MAX,MIN,NSCAL
  WRITE(3,3)
  RETURN
  END
```

FIG. F-8

```
MON$$     EXEQ LINKLOAD
          PHASEINCONT
          CALL INCONT
          PHASERAY
          CALL RAY
          PHASEMATINV
          CALL MATINV
          PHASEFORCST
          CALL FORCST
          PHASEPLOTTER
          CALL PLOTTER
MON$$     EXEQ INCONT.MJB
MON$$     EXEQ RAY.MJB
MON$$     EXEQ MATINV.MJB
MON$$     EXEQ FORCST.MJB
MON$$     EXEQ PLOTTER.MJB
```

FIG. 5-9

```
MON$$     JOB
MON$$     ASGN MJB,12
MON$$     ASGN MGO,16
MON$$     MODE GO,TEST
MON$$     EXEQ FORTRAN.......RAY
      INTEGER P
      DIMENSIONX(500),F(100),G(100),S(100)
      DIMENSIONC(100),W(100),R(100),A(100)
      DIMENSIONB(100),E(100),T(100)
  606 FORMAT(1X,12H SHEEP DATA )
    1 FORMAT(I10,F30.2)
    5 FORMAT(I3,I3)
      WRITE(3,606)
      WRITE(2,606)
      READ(1,5)N,M
C   CALL THE SUBROUTINE TO DETREND THE DATA
      CALL TREND(AX,X,N)
      SUM=0.
      DO 10I=1,N
   10 SUM=SUM+X(I)
      D=N
      AMEAN=(SUM)/D
      SX=0.0
      DO 3I=1,N
      X(I)=X(I)-AMEAN
    3 SX=SX+(X(I))**2
      SX=(SX/D)**.5
      DO 12I=1,N
   12 X(I)=X(I)/SX
      P=0
      TZ=0.0
      SZ=0.0
      DO 13I=1,N
      TZ=TZ+X(I)
   13 SZ=SZ+X(I)**2
      FZ=TZ
      GZ=SZ
      T(1)=TZ-X(1)
      F(1)=FZ-X(N)
      S(1)=SZ-X(1)**2
      G(1)=GZ-X(N)**2
      DO 14 P=2,M
```

FIG. E-10

```
      L=N-P+1
      K=P-1
      T(P)=T(K)-X(P)
      F(P)=F(K)-X(L)
      S(P)=S(K)-X(P)**2
   14 G(P)=G(K)-X(L)**2
      P=0
      DO 15 P=1,M
      C(P)=0.0
      J=N-P
      DO 16I=1,J
      MN=I+P
   16 C(P)=C(P)+X(I)*X(MN)
      H=J
      A(P)=H*C(P)-F(P)*T(P)
      B(P)=(H*G(P)-(F(P))**2)**.5
      E(P)=(H*S(P)-(T(P))**2)**.5
      R(P)=A(P)/(B(P)*E(P))
   15 WRITE(3,1)P,R(P)
      END
```

FIG. E-10 (CONTINUED)

```
MON$$       EXEQ FORTRAN,A.....TREND
   SUBROUTINE TREND(AX,X,N)
   DIMENSION X(500),AX(500)
11 FORMAT(23H THIS IS DETRENDED DATA)
 1 FORMAT(2F20.8)
 4 FORMAT(F20.8)
 6 FORMAT(F5.1)
   SUMY=0
   SSQY=0
   SUMXY=0
   SUMX=0
   SSQX=0
   DO50I=1,N
   READ(1,6)(AX(I))
   S=I
   SUMY=SUMY+S
   SSQY=SSQY+S**2
   SUMXY=SUMXY+AX(I)*S
   SUMX=SUMX+AX(I)
50 SSQX=SSQX+AX(I)**2
   R=N
   A1=(R*(SUMXY)-(SUMX)*(SUMY))/
   ((R)*(SSQY)-(SUMY**2))
   WRITE(3,4)(A1)
   WRITE(2,4)(A1)
   WRITE(3,11)
   WRITE(2,11)
   DO 60I=1,N
   S=I
   X(I)=AX(I)-(A1)*S
   WRITE(3,1)(X(I),AX(I))
60 WRITE(2,1)(X(I),AX(I))
   RETURN
   END
```

FIG. E-11

```
MON$$      JOB
MON$$      COMT 30 MINUTES. 4 PAGES
MON$$      ASGN MJB.12
MON$$      ASGN MGO.16
MON$$      MODE GO.TEST
MON$$      EXEQ FORTRAN.......RAY
      DIMENSIONX(500).A(100).B(100).R(100)
    1 FORMAT(I3)
    2 FORMAT(3F20.7.2I10)
   32 FORMAT(14X.F10.4)
      PI=3.1415927
      READ(1.1)ND
      TND=ND
      READ(1.1)LIM
      DO18I=1,ND
   18 READ(1.32)X(I)
      CONT=0.0
      DO3I=1.ND
    3 CONT=CONT+X(I)
      CONT=CONT/TND
C*****CALCULATE THE FOURIER COEFFICIENTS
      DO50N=5.LIM
      A(N)=0.0
      B(N)=0.0
      TOU=N
      NR=N
      MIS=N
   66 NR=NR+MIS
      IF(ND-NR)22,44.66
   22 NR=NR-MIS
   44 CONTINUE
      DO40I=1,NR
      F=I-1
      B(N)=B(N)+SIN(2.*PI*F/TOU)*X(I)
      A(N)=A(N)+COS(2.*PI*F/TOU)*X(I)
   40 CONTINUE
      ANR=NR
      B(N)=(2./ANR)*B(N)
      A(N)=(2./ANR)*A(N)
      R(N)=(A(N)**2+B(N)**2)**.5
      WRITE(3.2)A(N).B(N).R(N).N.NR
      WRITE(2.2)A(N).B(N).R(N).N.NR
   50 CONTINUE
      STOP
      END
```

FIG. F-12

THE APPLICATION OF GENERAL EXPONENTIAL SMOOTHING TO FORECASTING A TIME SERIES

by

RAYMOND CHARLES MILLER

B. S., University of Illinois, 1966

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

In many industrial and economic situations a series of discrete observations are taken on a uniform time scale. This sequence of observations is called a time series. The object of this investigation is to develop a method for forecasting these time series.

If the series of observations is not merely random it can be described as being the sum of two components. One component is the process which generates the time series and the other is the variation, or superimposed